Syntaxe théorique et formelle

Kim Gerdes Sylvain Kahane



No series description provided

ISSN: 2364-6209

Syntaxe théorique et formelle

Kim Gerdes Sylvain Kahane



Gerdes, Kim & Sylvain Kahane. 2020. *Syntaxe théorique et formelle* (Textbooks in Language Sciences). Berlin : Language Science Press.

This title can be downloaded at:

http://langsci-press.org/catalog/book/000

© 2020, Kim Gerdes & Sylvain Kahane

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

http://creativecommons.org/licenses/by/4.0/

ISBN: no digital ISBN

no print ISBNs!

ISSN: 2364-6209

no DOI

ID not assigned!

Cover and concept of design : Ulrike Harbort Fonts : Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: XALATEX

Language Science Press Xhain Grünberger Str. 16 10243 Berlin, Germany langsci-press.org

Storage and cataloguing done by FU Berlin



Av	ant-p	propos	ix
	1	De quoi parle ce livre?	ix
	2	Arbre de dépendance	xi
	3	Les questions qui nous préoccupent	xii
	4	À qui s'adresse cet ouvrage?	xiii
	5	La syntaxe et les autres domaines de la linguistique Q	xiv
	6	Grammaire et lexique	xvii
	7	Notations Q	xvii
	8	Le lexique : un cabinet de curiosités 🖟	xviii
	9	Le plan du livre	XX
	10	Les termes grammaire, syntaxe et topologie ②	xxii
	11	Commentaires sur le plan	xxiv
	12	Notions, termes, concepts et définitions Q	XXV
	13	Présentation de l'ouvrage	xxvii
	14	Remerciements	xxvii
I	Mo	odéliser la langue	3
1	La la	angue : L'objet d'étude de la linguistique	7
	1.1	Parler une langue	7
	1.2	La langue comme correspondance sens-texte 🕲	8
	1.3	Sons et textes	10
	1.4	Langue et variations 🚳	11
	1.5	Sens et intention communicative	13
	1.6	Mots et pensée @	15
	1.7	Sens lexicaux et traduction 🚳	15
	1.8	Langue, linguistique et modélisation	16
	1.9	Langue et parole, compétence et performance Q	17
	1.10	La planification	19
	1.11	Corpus et introspection	22
	1.12	Acceptabilité	23

	1.13	Parler et comprendre
2	Prod	luire un énoncé : La syntaxe mise en évidence par un exemple
	2.1	Analyse et synthèse
	2.2	Les observables : textes et sens ②
	2.3	Partir d'un sens
	2.4	Graphe et arbre 🕏
	2.5	Les composantes du sens Q
	2.6	Choisir des unités lexicales
	2.7	Les quatre moyens d'expression du langage Q
	2.8	Contraintes syntaxiques sur les choix lexicaux
	2.9	Règles et exceptions
	2.10	Structure hiérarchique
	2.11	Du sens au texte : de 3D à 1D Q
	2.12	Ce que la langue nous force à dire
	2.13	Les sens grammaticaux 🚳
	2.14	Choisir le verbe principal
	2.15	Les contraintes de la grammaire
	2.16	D'où viennent les règles de la grammaire ? 🕲
3	Lan	nodélisation : Préciser l'objectif de notre étude
•	3.1	Définition
	3.2	Modèle d'une langue ou modèle de la langue ②
	3.3	Modélisation et théorie
	0.0	3.3.1 Un exemple de modélisation en physique
		3.3.2 La modélisation de la langue
		3.3.3 Du modèle à la théorie
	3.4	Un exemple — l'accord — de la description à l'explication Q
	3.5	Modèle déclaratif
	3.6	Modèle génératif, équatif et transductif 🔮
	3.7	Modèle symbolique
	3.8	Calcul symbolique et grammaires catégorielles 🕏
	3.9	Modularité et stratification
	3.10	La Théorie Sens-Texte
	3.10	Modélisation des langues et ordinateur

II		s unités de la langue : Les trois composantes du signe lin- istique	77
4	Trois	s types d'unités : Morphème, syntaxème, sémantème	81
	4.1	Introduction	81
	4.2	Double articulation du langage	81
	4.3	Découverte de la double articulation 🕲	83
	4.4	Signifié, syntactique, signifiant	84
	4.5	Le signe linguistique Q	86
	4.6	Signifié ou exprimende Output	87
	4.7	Sémantème, syntaxème, morphème	87
	4.8	L'identification des unités de la langue	88
	4.9	Structuralisme et distributionnalisme 🕲	89
5	Mor	phèmes et syntaxèmes : Unités de forme vs unités de combinatoire	95
	5.1	Principe de commutation	95
	5.2	Commutation et exclusion mutuelle	96
	5.3	Décomposition propre des signes	97
	5.4	La quatrième proportionnelle Q	100
	5.5	Quasi-signes	102
	5.6	Combinaison libre	104
	5.7	Liberté de combinaison et opposition parole/langue 🍳	107
	5.8	Théorie des ensembles 🕏	108
	5.9	Dualité 🕏	109
	5.10	Signème	111
	5.11	Racine et signème	112
	5.12	Signèmes minimaux : morphème et syntaxème	113
	5.13	Les termes morphème et autres X -èmes \bigcirc	114
	5.14	Lexème, flexion et grammème	115
	5.15	Signème libérable, radical, affixe	116
	5.16	Dérivation et composition	117
	5.17	« Syntaxe » des morphèmes sous-lexicaux 🍳	118
	5.18	Sémantique des morphèmes sous-lexicaux Q	119
	5.19	À la limite entre morphème sous-lexical et syntaxème flexionnel	120
	5.20	Allomorphie	121
	5.21	Alternance vocalique en français : allomorphie ou alternance	
		phonologique?	124
	5.22	Amalgame, alternance et mégamorphe	124
	5.23	Flexions irrégulières Q	126

	5.24	Syntaxème zéro	126
	5.25	Syntaxème zéro et troncation en français 🖟	128
	5.26	Syntaxème et morphologie	129
6	Séma	antèmes et syntaxèmes : Unités de sens vs unités de combinatoire	141
	6.1	Arbitraire du sémantème	141
	6.2	Chaque sémantème suppose un choix	142
	6.3	Motivation du signe	144
	6.4	Syntaxème vide	146
	6.5	Constructions verbales et accords : signes vides ?	148
	6.6	Décomposition en sémantèmes	150
	6.7	Pronoms Q	151
	6.8	Phrasème	152
	6.9	Lexie et grammie	154
	6.10	Verbes supports et unités grammaticales 🝳	154
	6.11	Collocation et choix liés	155
	6.12	Fonctions lexicales	158
	6.13	Collocations morphologiques Q	159
	6.14	Délimitation paradigmatique des sémantèmes	159
	6.15	L'axe paradigmatique Q	161
	6.16	Dimension sémiotique du découpage	162
	6.17	Extensions paradigmatique et syntagmatique 🕏	163
	6.18	Syntaxème et faisceau de signes	164
	6.19	Signèmes et signes :unités de la langue et unités de la parole 🍳 👚	166
	6.20	Autonomie de la syntaxe et de la sémantique	167
III		uctures syntaxiques et topologique : Comment les unités	
	se o	combinent	175
7	-	axèmes et syntagmes : La délimitation des unités minimales de	
	la sy	ntaxe	179
	7.1	Syntaxème, morphème et sémantème	179
	7.2	Analogie structurelle	180
	7.3	Le collimateur et la sellette 🖟	181
	7.4	Syntagme ou syntaxème?	182
	7.5	À chacun son syntagme 🕲	183
	7.6	Syntaxe et morphologie	185
	7.7	Ce que la syntaxe n'est pas	186

	7.8	Historique de la notion de syntaxe 🕲	7
	7.9	Dimension paradigmatique du syntaxème	8
	7.10	Typologie des syntaxèmes	9
	7.11	Constructions N de N 18	9
	7.12	Lexème syntagmatique	0
	7.13	Constructions N+N ou N⊕N? ♠	1
	7.14	Construction syntagmatique isolée	3
	7.15	Inséparabilité linéaire	4
	7.16	Syntaxèmes discontinus 🚳	5
	7.17	Syntaxèmes séparables? 🚳	6
8	La c	onnexion syntaxique : Décrire les combinaisons entre unités	
	synt	axiques 20	5
	8.1	Structure syntaxique	5
	8.2	Justifier la structure syntaxique	6
	8.3	Faut-il définir la structure syntaxique? ②	7
	8.4	Nature de la structure	8
	8.5	Calculer la structure syntaxique d'un énoncé 🕲 20	9
	8.6	Structure syntaxique et structure topologique	1
	8.7	De la non-séparation des ordres au mouvement @ 21	2
	8.8	Connexion	3
	8.9	Unité syntaxique	5
	8.10	Test de commutation	7
	8.11	Unité syntaxique autonomisable	9
	8.12	Représenter les combinaisons	2
	8.13	Composition et décomposition	2
	8.14	La connexion et ses instances	4
	8.15	Définition formelle de la connexion 🕏	6
	8.16	Tests pour la connexion	6
	8.17	Connexion binaire? Q	9
	8.18	Structures de connexion, granularité et critères	1
	8.19	Structure de connexion et unités	2
	8.20	Les limites de la dualité Q	3
	8.21	Arbre de fragmentation	4
	8.22	De la fragmentation à la connexion	5
	8.23	Graphe à bulles et polygraphe 🕏	6
	8.24	Cycle de connexions Q	6
	8.25	Analyse en constituants immédiats	7

	8.26	Traitement cognitif des connexions	239
)	Tête	et dépendance : Hiérarchiser la structure syntaxique	249
	9.1	Hiérarchie	249
	9.2	Tête et dépendance	249
	9.3	Historique des notions de dépendance et de tête 🕲	252
	9.4	Dépendances morphologique, syntaxique et sémantique Q	260
	9.5	Structure de dépendance et arbre de dépendance	262
	9.6	Historique des représentations syntaxiques par des diagrammes en dépendance	263
	9.7	Arbre et représentation planaire	267
	9.8	Critère d'effacement simple et critère prosodique	268
	9.9	Tête d'un énoncé	271
	9.10	Tête d'un énoncé dans une langue inconnue 🚳	274
	9.11	Distribution et valence	275
	9.12	Le critère distributionnel avec effacement	278
	9.13	Tête effaçable	279
	9.14	Le critère distributionnel sans effacement	280
	9.15	Tête faible	282
	9.16	Quelle structure pour quels critères?	283
	9.17	Tête interne et critère rectionnel	284
	9.18	Critère de la tête sémantique	286
	9.19	Critère sémantique de non-effacement	287
	9.20	Distorsions sémantique-syntaxe Q	288
	9.21	Synthèse des critères pour la tête syntaxique	290
	9.22	Les cas problématiques	292
	9.23	« Déterminants complexes » 🖟	293
	9.24	Nom ou déterminant comme tête?	295
	9.25	Nom comme tête?	295
	9.26	Quand le dépendant contrôle son gouverneur 🍳	297
	9.27	Déterminant comme tête?	298
	9.28	Nom et déterminant comme co-têtes	302
	9.29	Co-occupation Q	302
	9.30	Les grammèmes comme tête Q	305
	9.31	Dominance et projections maximales	305
	9.32	Dépendance, dominance et transitivité 🕏	307
	9.33	Critère de recouvrement	309
	9.34	Multiplicité des analyses en dépendance	310

	9.35	Dépendance, constituance et topologie	312
10	Cons	stituance et dépendance : Différentes représentations de la com-	
	binai	ison	333
	10.1	De la dépendance à la constituance	333
	10.2	Arbre à la Beauzée-Gladkij	333
	10.3	Arbre de constituants plat	335
	10.4	Le terme constituant 🕲	336
	10.5	Arbre de constituants avec têtes	337
	10.6	Equivalence par la méthode de Lecerf	337
	10.7	Construire un arbre de dépendance par décomposition	338
	10.8	Décomposition récursive 🕏	339
		Décomposition vs. combinaison	340
		Les limites des méthodes par décomposition Q	341
		Les tests de constituance	342
		.	344
		Le statut des tests de constituance Q	346
		Pas de structure de constituants sans la notion de tête? Q	349
		Arbre de constituants binaire	350
		Projections partielles et ordre de saillance	351
		Dépendance et constituance se complètent Q	352
	10.18	Historique des représentations syntaxiques par des diagrammes	
		en constituants 🖲	353
		Binarité, construction et connexion Q	363
	10.20	Syntaxe X-barre 🍳	364
		Le groupe verbal ②	366
	10.22	Arbre de constituants binaire et polygraphe	367
		Réification et transitivité 🕏	368
		Polygraphe orienté et arbre de dépendance	369
		Comparaison des méthodes par décomposition	370
	10.26	Équivalences entre structures	371
		Deux types d'arbres de constituants Q	371
		Combiner les méthodes ascendante et descendante	374
	10.29	Enseigner la syntaxe	377
	10.30	Constituance ou dépendance?	379
11		pologie : Ordre des mots, linéarisation et regroupements	391
	11.1	La linéarité de la langue	391
	11.2	Les exceptions à la segmentabilité	392

11.3	Ordre des mots vs. topologie	393
11.4	Place linéaire et position syntaxique	394
11.5	Position topologique	395
11.6	Le cas des langues V2 🚳	396
11.7	Linéarisation	398
11.8	Une représentation syntaxique non ordonnée ? Q	399
11.9	Mouvement et ordre de base Q	400
11.10	Arbre de dépendance ordonné	402
		403
		404
11.13	Langues à têtes finales et langues à têtes initiales 🏵	406
11.14		408
11.15	Projectivité	409
	Projectivité et planarité 🕏	411
11.17	Équivalence des définitions de la projectivité 🍳	412
11.18	Flux de dépendances	413
	Linéarisation projective	415
11.20	Linéarisation des co-dépendants	417
11.21	Un cas d'exclusion étrange Q	419
11.22	Préférences dans l'ordre	420
11.23	Contraintes de poids ②	422
11.24	Les langues dites à ordre libre 🚳	423
11.25	Ordre communicativement dirigé 🚳	425
11.26	Les autres usages de l'ordre linéaire 🔍	427
	Modèle topologique	428
11.28	Topologie du groupe substantival en français 🖟	430
11.29	Arbre de constituants ordonné 🕏	431
11.30	Grammaire de réécriture 🕏	432
11.31	Projection topologique	434
11.32	Émancipation	436
	Non-projectivité en français	436
11.34	Grammaire topologique formelle 🕏	438
	Constituants topologiques intermédiaires	441
11.36	Topologie et prosodie	442
11.37	La structure topologique de l'allemand 🚳	443
	L'anglais comme langue de référence	445
11.39	Historique de la description topologique 🕲	445

Avant-propos

DE LA SYNTAXE *Ou la manière de joindre ensemble les parties d'oraison* [= les parties du discours ou catégories lexicales] *selon leurs divers régimes*.

Ces diverses parties font pour ainsi dire par rapport à une langue, ce que font les matériaux, par rapport à un édifice : quelque bien préparés qu'ils soient ils ne feront jamais un palais ou une maison, si on ne les place conformément aux règles de l'architecture. C'est donc la syntaxe qui donne la forme au langage, et c'est la partie la plus essentielle de la grammaire.

(Claude Buffier, *Grammaire françoise sur un plan nouveau*, **1709**, p. 294, nous modernisons l'orthographe)

1 De quoi parle ce livre?

En commençant cet ouvrage, nous souhaitions écrire un ouvrage d'introduction à la SYNTAXE DE DÉPENDANCE. La plupart des ouvrages récents en syntaxe s'appuient sur l'ANALYSE EN CONSTITUANTS qui a dominé la seconde moitié du 20^e siècle. Et si les grammaires de dépendance ont connu un renouveau et un développement extraordinaire depuis le début du 21^e siècle, jusqu'à supplanter quasi totalement les grammaires de constituants dans le domaine du traitement automatique des langues (TAL), elles ne sont encore que sporadiquement enseignées à l'université et l'unique ouvrage de référence reste souvent l'ouvrage fondateur de Lucien Tesnière publié en Tesnière1959. Les Éléments de syntaxe structurale de Tesnière sont un incontestable monument de la littérature scientifique en linguistique, dont on ne peut que recommander la lecture, mais cet ouvrage ne peut évidemment pas prendre en compte les développements important qu'a connu le domaine depuis 60 ou 80 ans. (Tesnière est mort en 1954, il a été très malade après la guerre et ses idées ont peu évolué depuis l'édition de son polycopié Esquisses de syntaxe structurale distribué aux élèves de l'École normale d'institutrices de Montpellier en 1943, voire de son article de 1934 Comment construire une syntaxe.)

Au final, le livre que vous avez entre les mains n'est pas un manuel sur la syntaxe de dépendance, dans le sens où il **ne** souhaite **pas** livrer de recettes qui per-

mettront au lecteur d'apprendre à associer un arbre de dépendance à n'importe quel énoncé et de discuter les différentes analyses possibles d'une construction donnée dans un cadre préconçu. L'objectif de ce livre est au contraire de s'interroger sur le cadre lui-même, de mettre en question la validité d'une approche de la syntaxe en termes de dépendance et au-delà de cela de définir les principes mêmes qui doivent présider à une construction théorique en syntaxe.

Il en découle que cet ouvrage n'est pas vraiment un ouvrage d'introduction : même s'il tente d'élaborer son objet à partir de rien et qu'il est donc en théorie accessible sans pré-requis, cet ouvrage fait certainement appel à une maîtrise du raisonnement scientifique qui ne s'acquière qu'avec l'expérience. On pourra comparer, toutes choses égales par ailleurs, une tentative de ce genre à celle du groupe de mathématiciens français rassemblés sous le pseudonyme de Nicolas Bourbaki qui rédigea un ouvrage de construction des mathématiques à partir de rien (le premier tome démarre par la construction des entiers à partir du seul ensemble vide). L'ouvrage de Nicolas Bourbaki, s'il est élémentaire au sens premier du terme (comme le souligne son titre, Éléments de mathématiques), s'avère d'une lecture bien difficile pour des non-mathématiciens.

Ce livre n'est pas non plus vraiment un ouvrage d'introduction à la syntaxe de dépendance, puisqu'une grande partie de l'ouvrage est consacrée à **définir l'objet même d'un ouvrage de syntaxe** et que la dépendance n'y est introduite qu'après une discussion détaillée sur les **unités de base de la syntaxe**. De plus, une large place est faite aux autres **représentations possibles de l'organisation syntaxique** et à la comparaison entre les différents modes de représentation. Nous pensons notamment que ceux qui travaillent en syntaxe de constituants en apprendront beaucoup sur les représentations qu'ils ont l'habitude d'utiliser et sur les **choix qui président à de telles représentations**.

Nous allons préciser l'objectif de cet ouvrage. Avant cela, le lecteur qui n'est pas familier avec la notion d'arbre de dépendance pourra consulter l'encadré qui suit. Dans la suite de l'ouvrage nous mettrons souvent des portions de texte en exergue de cette façon.



Q

2 Arbre de dépendance **Q**

Tout au long de cet ouvrage, nous proposerons de petits encadrés. Certains anticipent un peu sur la suite de l'ouvrage; celui-ci permet à un lecteur totalement néophyte d'avoir une première idée de ce qu'est la SYNTAXE DE DÉPENDANCE. De manière générale, les encadrés contiennent des informations complémentaires, généralement plus techniques ou à visée historique, qui ne sont pas essentielles à la compréhension du texte principal.

L'arbre de dépendance est une représentation de la structure syntaxique devenue traditionnelle après la publication en 1959 de l'ouvrage de Lucien Tesnière, Éléments de syntaxe structurale, et les différents travaux qui ont suivi, notamment ceux des pragois autour de Petr Sgall, ceux des Russes autour d'Igor Mel'čuk, ainsi que des travaux en Allemagne, en Angleterre ou aux. États-Unis (mais étonnamment aucun travail significatif en France jusqu'aux années 1990).

Dès le début de son ouvrage, Tesnière dit :

« Tout mot qui fait partie d'une phrase cesse par lui-même d'être isolé comme dans le dictionnaire. Entre lui et ses voisins, l'esprit aperçoit des **connexions**, dont l'ensemble forme la charpente de la phrase. »

(Voir l'encadré ?? pour un *Historique des notions de dépendance et de tête*, où l'on verra que cette idée est déjà dans un article de l'*Encyclopédie* par Dumarsais en 1754 et que Tesnière a eu de nombreux prédécesseurs.) Tesnière ajoute ensuite que ces connexions sont orientées, liant un gouverneur à un dépendant, et forment ainsi une structure hiérarchique. C'est ce que Tesnière appelait un STEMMA et qu'on appelle aujourd'hui un ARBRE DE DÉPENDANCE (voir l'encadré ?? sur *Graphe et arbre*). Ainsi l'arbre de dépendance généralement proposé pour une phrase comme :

(1) Beaucoup de gens aimeraient passer Noël en Laponie.

est:

Dans cette représentation, les mots dépendent les uns des autres. Le mot le plus important, le verbe principal, est *aimeraient*, qui occupe la racine de l'arbre et est placé tout en haut (l'arbre « pousse » à l'envers). Les dépendances sont étiquetées par des relations syntaxiques. Ainsi l'arbre nous dit que le sujet de *aimeraient* est le groupe *beaucoup de gens* et que le mot le plus important de ce groupe est *beaucoup*, qui se retrouve ainsi lié à *aimeraient*.

Nous ne justifierons pas ici cette représentation, dont on peut d'ailleurs contester certains choix. Cela sera très largement discuté dans cet ouvrage et en particulier dans le Chapitre 9 sur *Tête et dépendance*. Il s'agit juste de donner un premier exemple en vue de la discussion qui va suivre.

3 Les questions qui nous préoccupent

En écrivant cet ouvrage, nous nous sommes posé un grand nombre de questions auxquelles il nous a semblé qu'il fallait répondre avant de pouvoir présenter de façon objective la syntaxe de dépendance.

- Est-il justifié de représenter la structure syntaxique par un arbre de dépendance? Par une structure arborescente? Par des dépendances? Jusqu'à quel point une représentation basée sur la dépendance est-elle ou non équivalente à d'autres modes de représentation et notamment aux arbres de constituants?
- Quel est le statut de la structure syntaxique? Est-ce un objet de la langue ou bien un artefact de la modélisation des langues? Quel rôle souhaite-t-on donner à de telles structures à l'intérieur du modèle d'une langue?
- Les dépendances syntaxiques sont-elles entre les mots? Quelles sont les unités minimales de la syntaxe? Comment définir le mot et quel rôle jouet-il dans la syntaxe s'il n'est pas l'unité minimale de la syntaxe?
- Les dépendances syntaxiques s'arrêtent-elles à la frontière de la phrase? La notion de phrase est-elle légitime? Y a-t-il une unité maximale de la syntaxe?
- De quelles propriétés des énoncés cherche-t-on à rendre compte par un arbre de dépendance? De quelles propriétés ne rend-on pas compte? Comment encoder les propriétés qui ne sont pas prise en compte par l'arbre de dépendance? Quelles sont les autres structures que l'on peut associer à un énoncé? Quels rapports y a-t-il entre les différentes représentations de la structure d'un énoncé?
- Finalement, qu'appelle-t-on la syntaxe? Et en quoi est-il possible ou non, nécessaire ou non, d'introduire une structure syntaxique pour rendre compte des propriétés syntaxiques d'un énoncé? La modélisation du lien entre signifiant et signifié a-t-elle besoin d'une structure syntaxique intermédiaire?

Toutes ces questions nous amèneront à commencer par rappeler les objectifs de notre discipline, la linguistique, et de sa sous-discipline, la syntaxe. Ces objectifs sont pour nous de **construire des modèles** des différentes langues du monde au sein d'un théorie de la langue. Notre première partie sera donc consacrée à définir les objectifs de la modélisation des langues et à montrer l'existence d'un ensemble de propriétés qui relèvent de ce que nous appelons la syntaxe.

On définit traditionnellement la syntaxe comme « l'étude de l'organisation des mots dans la phrase ». Une telle définition est problématique, puisqu'elle suppose que l'on peut définir les notions de mot et de phrase avant de définir ce qu'est la syntaxe. Dans cet ouvrage, la notion de mot ne sera définie qu'au début de la quatrième partie et l'unité maximale de la syntaxe ne sera discutée que dans la sixième et dernière partie de l'ouvrage.

Avant d'étudier l'organisation syntaxique, nous tenterons de **caractériser la syntaxe** et notamment les unités minimales de la syntaxe, que nous contrasterons avec les unités minimales de la morphologie et de la sémantique. Ce sera notre deuxième partie.

La troisième partie montrera comment définir une structure qui rend compte des principales propriétés syntaxiques des énoncés. Différentes structures seront présentées et la représentation par un arbre de dépendance sera particulièrement discutée.

Les trois parties suivantes s'apparentent davantage à un manuel traditionnel. Nous y présenterons différentes caractéristiques des langues et notamment du français et nous présenterons les structures qui en rendent le mieux compte. On notera néanmoins que les catégories syntaxiques et autres parties du discours, qui sont généralement introduites très tôt dans les ouvrages de syntaxe, ne seront réellement définies que dans les quatrièmes et cinquièmes parties, quand la question de l'organisation des unités aura été largement discutée.

Nous précisons le plan de cet ouvrage dans la section 9.

4 À qui s'adresse cet ouvrage?

Cet ouvrage aborde la syntaxe comme une composante d'un modèle linguistique. L'idée même que la langue puisse être modélisée, comme peut l'être le mouvement des planètes ou le développement du fœtus, n'est pas nécessairement acceptée par tous ceux qui s'intéressent aux langues et prennent du plaisir à les apprendre ou les étudier. Cet ouvrage souhaite montrer qu'on peut dégager de manière méthodique les propriétés des langues, mettre de l'ordre dans la forêt vierge que constitue chaque langue et élaborer un objet théorique qui reproduise certaines propriétés d'un locuteur qui parle et que nous appelons un modèle d'une langue.

Il existe en linguistique, comme ailleurs en sciences humaines, des courants théoriques variés, plus ou moins d'accord entre eux. Cet ouvrage ne se situe pas précisément dans un courant dominant en linguistique, mais il puise largement dans le courant structuraliste qui s'est développé depuis un siècle, des travaux

pionniers de Ferdinand de Saussure aux travaux actuels en linguistique formelle, en passant par les travaux précurseurs d'Otto Jespersen, ceux des distributionnalistes américains, Leonard Bloomfield en tête, et ceux de Lucien Tesnière et d'Igor Mel'čuk en syntaxe de dépendance. Les auteurs de ce livre ont une formation initiale en mathématiques et ont travaillé dans le domaine du traitement automatique des langues, des grammaires formelles et de la modélisation mathématique des langues et ils enseignent la linguistique à tous les niveaux universitaires. Bien que cet ouvrage ne traite pas directement de formalisation mathématique et d'implémentation informatique, il se place dans le cadre d'une approche déductive de la langue dont l'objectif est de construire des modèles qui peuvent être formalisés et implémentés pour simuler un locuteur humain. Ce n'est néanmoins pas l'objectif de ce livre de présenter des modèles de la langue; ce livre se contente d'introduire, de la façon la plus rigoureuse possible, les notions nécessaires à l'étude de la syntaxe, en se concentrant sur les structures syntaxiques et non sur les règles de la grammaire.

Cet ouvrage a une visée à la fois scientifique et pédagogique : il a été élaboré avec l'objectif de fournir une base pour l'enseignement de la syntaxe à l'université et de présenter des notions fondamentales pour l'étude des langues, aussi diverses soient-elles. Nous espérons qu'il montre qu'il est utile d'enseigner la syntaxe pour comprendre le fonctionnement des langues et mieux les enseigner et les apprendre. Cet ouvrage souhaite également montrer que la syntaxe est un domaine de recherche vivant et que de nombreuses questions restent ouvertes, même pour des langues très étudiées comme le français.

Cet ouvrage écrit en français est évidemment destiné aux francophones et à ceux qui apprennent le français. Le français constituera donc la langue que nous étudierons par défaut. Bien que cet ouvrage ne soit pas une grammaire méthodique du français, il constitue une bonne introduction à la syntaxe du français. Mais il souhaite aussi fournir les outils nécessaires à l'étude d'autres langues, même très éloignées du français. À chaque fois que cela sera utile, nous montrerons le caractère exotique du français dans la diversité des langues et présenterons pour la notion étudiée un fonctionnement différent de celui du français dans une autre langue.

5 La syntaxe et les autres domaines de la linguistique 🍳



La syntaxe n'est qu'une partie d'un modèle linguistique, bien que dans beaucoup de théories linguistiques, elle occupe une place centrale ou même dominante par rapport aux autres domaines. Avant de préciser ce qu'est la syntaxe,



nous allons situer celle-ci parmi les autres domaines de la linguistique. Nous allons le faire en répondant à différentes questions que l'on peut se poser concernant la langue.

Comment fonctionne l'esprit? Comment fonctionne le raisonnement? Comment les modéliser, les imiter?

Ce genre de questions appartient à la psychologie, la neurologie, la logique, l'intelligence artificielle, mais on peut espérer que de nouvelles connaissances en linguistique éclaireront aussi ces sujets. En quelque sorte, la linguistique peut être considérée comme un sous-domaine de chacune des disciplines citées : elle couvre les questions liées à la langue qui se posent dans ces sciences. Parfois, on classe toute cette thématique sous le terme de sciences cognitives.

Que veut dire le dernier énoncé que j'ai entendu? Comment ai-je pu en extraire le sens? Comment combiner le sens des mots pour former le sens des énoncés? Ou en quoi la phrase « Paul et Marie porte un chapeau » se distingue-t-elle de la phrase « Paul et Marie porte une machine à laver »?

Ces questions caractérisent la sémantique. Même si elles ne relèvent pas directement de la syntaxe, nous les aborderons à plusieurs reprises, pour mieux délimiter la syntaxe, mais aussi parce que la syntaxe s'articule directement avec la sémantique.

Pourquoi dit-on telle ou telle phrase dans un contexte particulier? Quelle est la signification d'un énoncé dans un contexte? Dans des termes plus techniques : quel est le but de l'acte de langage et comment est-il poursuivi? Concrètement : pourquoi doit-on donner à la question « Vous avez l'heure? » la réponse « Il est trois heures » et pas la réponse « Oui »?

Ces questions font partie de la pragmatique et se situent au-delà de la sémantique. Nous n'y toucherons pas.

Dans quel ordre doit-on placer les mots? Quand doit-on utiliser un verbe ou un nom? A-t-on le droit de coordonner des pronoms interrogatifs? Quel type de structure forment les mots assemblés en une phrase?

Celles-ci et beaucoup d'autres questions que nous étudions dans ce livre concernent la syntaxe.

Comment représenter et récupérer l'information sur le fonctionnement de chaque mot? Pourquoi dit-on une peur bleue, mais vert de peur?

Ces informations sont dans le lexique. La lexicologie est la science qui étudie le lexique. Nous aborderons ces questions dans la mesure où les particularités lexicales ont une influence sur la syntaxe et où la frontière entre les constructions syntaxiques et le lexique proprement dit est plus que mouvante. La question est discutée avec plus de détails dans la section suivante.

Comment sont formés les mots? De quelles entités sont-ils formés? Pourquoi peut-on dire désespéré et non désattendu? Comment se forme la 2^e personne singulier du passé simple pour le verbe aimer?

La **morphologie** tente de répondre à ces questions. Nous verrons qu'une partie de ces questions relèvent pour nous de la syntaxe (par exemple la conjugaison des verbes).

Quels sont les sons d'une langue? Quelles sont les combinaisons possibles de ces sons? Par exemple, pourquoi les Espagnols mettent-ils des e devant chaque mot dont l'équivalent français commence par sp comme especial? Pourquoi les Français prononcent-ils ze au lieu de the quand ils parlent anglais? Dans la phrase « Le fromage je n'aime vraiment pas », pourquoi la mélodie monte-t-elle sur le mot fromage? Comment est représentée la prononciation d'un mot dans le cerveau?

La phonologie s'interroge sur ces questions. Nous les aborderons dans la mesure où la structure phonologique des énoncés nous offre de nombreux indices sur la structure syntaxique des énoncés et où la langue parlée constitue, du notre point de vue, un bien meilleur sujet d'étude que la langue écrite.

Comment sont réellement formés les sons de la langue ? Comment se distingue un son d d'un son z? Pourquoi y a-t-il une différence entre un son k prononcé devant i et un son k prononcé devant u? Comment se distingue le son a de l'allemand du son a du français.

S'il s'agit de répondre à ces questions en termes de fonctionnement de la langue et de l'humain produisant les sons, on se trouve dans la **phonétique** (articulatoire), si c'est en termes purement techniques et physiques, c'est plutôt l'acoustique et le traitement du signal, un domaine relevant de la physique, qui sont concernés.

Il existe d'autres domaines en linguistique, notamment tout ce qui concerne la langue dans ses variations dans le temps et dans l'espace : sociolinguistique, géolinguistique, dialectologie, diachronie, origine du langage, génétique des langues, grammaticalisation, créolisation, productivité lexicale, acquisition de la langue par l'enfant, apprentissage d'une langue seconde. Il existe aussi de nombreux domaines d'application : le traitement automatique des langues ou TAL (traduction automatique, recherche d'information sur le web ou dans des bases de données, aide aux handicapés, synthèse de la parole), l'enseignement de la langue (Français Langue Étrangère ou FLE pour les apprenants langue seconde, la didactique des langues pour les écoliers francophones), etc. Dans les champs interdisciplinaires, nous trouvons la psycholinguistique, la sociolinguistique, la neurolinguistique, la linguistique textuelle, la linguistique mathématique, etc.

6 Grammaire et lexique

Lorsqu'un locuteur veut énoncer une idée dans sa langue, il doit trouver dans son lexique les unités lexicales qui correspondent le mieux aux entités dont il veut parler. Mais, pour former la phrase, le locuteur a besoin d'autres éléments linguistiques qui sont contraints de diverses façons par la langue. L'étude de ces éléments et des contraintes que la langue impose à un locuteur s'appelle la GRAM-MAIRE de cette langue.

En fait, la frontière entre lexique et grammaire n'a rien d'évident. La description d'une langue est la description de chacune des unités de la langue et de la façon dont elles se combinent. Parmi ces unités, on trouve des unités lexicales prototypiques comme CHEVAL ou MANGER, tandis que d'autres sont des unités grammaticales incontestables comme le temps imparfait (Le cheval mangeait) ou les différentes réalisations syntaxiques du pluriel (Les chevaux mangeaient). Mais d'autres unités, tout en partageant des propriétés avec les unités lexicales prototypiques ont aussi un fonctionnement grammatical, comme chose ou faire (La chose que je préfère, c'est manger; Ce que je préfère faire, c'est manger). Inversement, des unités qui ont un fonctionnement a priori grammatical, comme la préposition à dans Elle parle à Pierre, auront un fonctionnement beaucoup plus lexical lorsqu'elles commutent avec d'autres unités : Elle est à la maison, dans la maison, devant la maison, derrière la maison, etc.

La distinction entre lexique est grammaire est orthogonale à la partition du modèle linguistique entre morphologie, syntaxe et sémantique. Toutes les unités, qu'elles soient lexicales ou grammaticales, possèdent une forme, un sens et une combinatoire qu'il faut décrire (voir la section 4.4 sur Signifié, signifiant, syntactique). La syntaxe est l'étude de la combinatoire des unités lexicales et grammaticales et tout particulièrement des combinaisons libres obéissant à des règles générales. La syntaxe se trouve à mi-chemin de la sémantique qui s'intéresse au sens des unités lexicales et grammaticales et des énoncés qu'elles forment en se combinant et de la MORPHOLOGIE qui s'intéresse à la forme et la structure des unités que la syntaxe combine.

7 Notations **Q**



Nous notons nos exemples linguistiques en italiques. Pour les unités lexi-CALES, nous utilisons des petites capitales. Lorsqu'il s'agit d'unités lexicales multimots comme [□]POMME DE TERRE , nous utilisons des balises angulaires. Les unités grammaticales sont quant à elles généralement désignées par des termes métalin-



guistiques : présent, singulier, féminin, etc. Le sens d'une unité lexicale ou d'une portion de texte est indiqué en guillemets simples : 'cheval', 'pomme de terre', 'le cheval mange'. La signification d'une unité grammaticale est également noté entre guillemets, mais en mettant le terme en petites capitales : 'SINGULIER'. On ne confondra pas le sens grammatical 'PRÉSENT' qui signifie 'ayant lieu maintenant' avec le sens lexical 'présent'.

8 Le lexique : un cabinet de curiosités 🖟

Bien que la syntaxe et la grammaire soient au centre de cet ouvrage, on ne peut pas ne pas évoquer la complexité lexicale dans un tel ouvrage. Nous allons en donner trois exemples.

Chaque verbe impose à ses compléments une construction particulière : manger quelque chose, parler \dot{a} quelqu'un de quelque chose, donner quelque chose \dot{a} quelqu'un, compter sur quelqu'un, aller quelque part, poser quelque chose quelque part, etc. Ces constructions se comptent en dizaines. Même lorsque ces constructions semblent similaires, comme parler à quelqu'un et penser à quelqu'un, elles peuvent différer par leur comportement : ainsi, à Marie, je lui parle, j'y pense ou je pense à elle, mais on ne pourra pas dire *je lui pense ou ?? je parle à elle (pour l'utilisation des symboles * et ??, voir la section 1.12 sur l'*Acceptabilité*). Dans certains cas, le verbe contraint tellement son complément que seules quelques formes sont acceptables. C'est le cas par exemple de la tournure verbale y comprendre quelque chose, qui n'est possible qu'avec les compléments suivants : Je n'y comprends rien, Je n'y comprends pas grand-chose, Que puis-je y comprendre?, Y comprends-tu quelque chose? et J'y comprends que dalle. Il est impossible d'avoir un groupe nominal référentiel comme complément : *7'y comprends une chose intéressante. La liste des compléments possibles de cette acception de COM-PRENDRE constitue ainsi un véritable cabinet de curiosités avec un pronom interrogatif (QUOI et sa forme atone que), un pronom négatif (RIEN), deux pronoms indéfinis — [□]QUELQUE CHOSE [□] qui n'est possible ici qu'avec l'interrogation et grand-chose qui est toujours accompagné de la négation — et enfin [∞]oue DALLE. Notons que d'autres tournures verbales possèdent quasiment la même complémentation : Ça ne rime à rien, Ça ne rime pas à grand-chose, À quoi ça rime?, mais pas *Ça rime à une chose intéressante ou *Ça rime à faire ça.

Les exceptions lexicales sont encore plus nombreuses quand on se rapproche de la grammaire. Le français possède par exemple plusieurs éléments négatifs qui se construisent avec ne : Je ne dors pas, Je ne dors plus, Je ne dors jamais, Je ne dors nulle part, Je ne mange rien, Je ne parle à personne, Je n'ai aucun



problème, Je n'ai qu'une idée. Chacun de ces éléments possède des propriétés syntaxiques différentes : par exemple JAMAIS peut être déplacé, mais pas PAS ou PLUS : Jamais je ne dors vs *Plus je ne dors. JAMAIS et PLUS peuvent être combinés, mais pas JAMAIS et PAS : Jamais plus je ne dormirai, Je ne dormirai plus jamais vs *Je ne dormirai pas jamais, *Je ne dormirai jamais pas. Notons encore que RIEN et PERSONNE se placent différemment par rapport au verbe : Je n'ai vu personne, Je n'ai rien vu. Sans aller plus loin, on aura compris que chacun de ces éléments négatifs nécessitera une étude séparée simplement pour déterminer ses propriétés combinatoires, c'est-à-dire sa « syntaxe ». Il en va de même de chacun des pronoms interrogatifs ou de chacun des pronoms relatifs et ainsi de la plupart des unités lexicales ayant un rôle grammatical.

Terminons par l'exemple des constructions, ainsi que l'on nomme les configurations qui possèdent un rôle grammatical. Il existe en français une construction très employée, le présentatif $^{\boxtimes}$ IL y A ... QU- $^{\boxtimes}$, pratiquement obligatoire à l'oral lorsque le sujet est indéfini :

- (2) Il y a quelqu'un qui nous regarde depuis la fenêtre.
- (3) Il y a des choses que j'ai achetées là-bas sur la table.
 Le présentatif peut être combiné avec la restriction en ^図NE ... QUE :
- (4) Il n'y a que des choses que j'ai achetées là-bas sur la table.

La combinaison du présentatif avec la restriction s'applique à des compléments indirects :

- (5) Il n'y a qu'à un endroit qu'on les trouve.
- (6) Il n'y a qu'à elle que je pense.

alors que le présentatif seul ne le peut pas :

- (7) *Il y a à un endroit qu'on les trouve.
- (8) *Il y a à quelqu'un que je pense.

Lorsque le présentatif s'applique à un complément avec possessif, on peut exprimer cette possession dans la forme même du présentatif :

(9) Il y a mon frère qui doit venir.

(10) J'ai mon frère qui doit venir.

Enfin, la même configuration peut être utilisée pour introduire un complément de temps avec le sens 'depuis' :

(11) Il y a une semaine qu'on ne s'est pas vu.

Dans ce cas, elle possède une variante, mais celle-ci n'est possible que pour les compléments de temps :

- (12) Ça fait une semaine qu'on ne s'est pas vu.
- (13) *Ça fait des choses que j'ai achetées là-bas.

Comme on le voit, on retrouve pour ces constructions de nombreuses **idiosyncrasies** qui justifient de leur donner une description détaillée, au même titre que les autres unités lexicales.

9 Le plan du livre

Ce livre est divisé en six parties que nous avons esquissées à la section 3 sur *Les questions qui nous préoccupent.* Nous allons préciser le plan du livre.

La première partie explique en quoi consiste une LANGUE et la MODÉLISATION de cette langue (Chapitre 1) et quelles sont les caractéristiques du modèle linguistique que nous construisons (Chapitre 3). Cette première partie permet donc de comprendre quel est le cadre théorique de cet ouvrage et avec quel objectif nous souhaitons mener notre étude de la langue et de sa syntaxe. La modélisation est illustrée par l'exemple de la production d'un énoncé (Chapitre 2).

La deuxième partie pose la question des unités minimales de la langue (Chapitre 4). Étudier la combinatoire des unités qui constituent les énoncés ne peut se faire qu'après avoir identifié les unités qui se combinent et notamment les unités minimales. Nous montrons qu'il est nécessaire de considérer trois types d'unités minimales : les MORPHÈMES ou unités minimales de forme (Chapitre 5), les sémantèmes ou unités minimales de sens (Chapitre 6) et les syntaxèmes ou unités minimales de la syntaxe, c'est-à-dire de la combinatoire libre. La distinction de trois types d'unités résulte de la non-correspondance entre les unités de forme et de sens. Prenons un exemple :

(14) Les étudiants m'ont donné un coup de main.

Il y a dans cet énoncé plusieurs sémantèmes qui sont exprimés par une combinaison de morphèmes : *coup de main* bien sûr, qui ne signifie pas ici un coup de la main, mais aussi *étudiant*, qui combine le radical du verbe ÉTUDIER avec le morphème -*ant* ou encore le passé composé exprimé par le verbe AVOIR combiné avec le morphème de participe passé -*é*.

La troisième partie introduit les unités syntaxiques et la façon dont celles-ci se combinent pour former la STRUCTURE SYNTAXIQUE. On y définit la syntaxe comme l'étude des combinaisons libres d'unités et on caractérise plus précisément le syntaxème (Chapitre 7). Nous montrons que les différentes fragmentations d'un énoncé en unités syntaxiques définissent un graphe que nous appelons la STRUCTURE DE CONNEXION et qui décrit les combinaisons entre syntaxèmes (Chapitre 8). On peut en plus hiérarchiser cette structure en considérant la notion de tête d'une unité syntaxique et obtenir ainsi une structure de dépendance (Chapitre 9). On montre comment représenter cette structure de manière plus ou moins équivalente par un Arbre de constituants (Chapitre 10). On s'intéresse ensuite au lien d'une part entre la structure et le texte et d'autre part entre la structure syntaxique et le sens. Le premier cas concerne l' « ordre des mots », c'est-à-dire à la façon dont les syntaxèmes s'ordonnent les uns par rapport aux autres et se regroupent pour former des constituants topologiques au sein de la structure topologique (Chapitre 11). Le deuxième cas concerne l'interface SÉMANTIQUE-SYNTAXE, c'est-à-dire à la façon dont les sémantèmes se combinent, ce que décrit la STRUCTURE SYNTAXIQUE PROFONDE, qui rend compte de la distinction entre actant et modifieur et des restructurations parfois complexes entre la représentation sémantique et la structure syntaxique (Chapitre ??).

Les trois parties suivantes présentent les trois grands domaines de la syntaxe, que nous appelons nanosyntaxe, microsyntaxe et macrosyntaxe, et les principales notions de la syntaxe seront présentées : le mot, les catégories flexionnelles, les catégories lexicales, les fonctions syntaxiques, la phrase. Les principales constructions seront également étudiées, et notamment les listes, l'extraction et l'organisation des énoncés autour d'un noyau.

La quatrième partie de ce livre est donc consacrée à la nanosyntaxe ou morphosyntaxe. Elle présente les combinaisons de syntaxèmes possédant une très grande cohésion, dont les composantes sont indissociables et se situent à l'intérieur ou à la frontière des mots. Elle inclut la syntaxe flexionnelle et la syntaxe des particules, c'est-à-dire la syntaxe de tous les éléments qui sont des marqueurs grammaticaux et qui possèdent très peu d'indépendance syntaxique. Nous montrons en particulier que le mot n'est qu'un degré particulier dans l'échelle de cohésion des combinaisons de syntaxèmes en unités syntaxèmes en un

taxiques, même s'il constitue une unité naturelle et l'unité qui a été privilégiée pour la transcription écrite de nombreuses langues (Chapitre ??). Nous terminons cette partie par une première classification des unités minimales de la syntaxe et introduirons les catégories flexionnelles (Chapitre ??) et les catégo-RIES NANOSYNTAXIQUES de LEXÈMES (Chapitre ??).

La cinquième partie est consacrée à la MICROSYNTAXE, c'est-à-dire la syntaxe de rection : la RECTION se caractérise par une relation hiérarchique avec des contraintes de réalisation imposées par un gouverneur à ses dépendants. Elle constitue la syntaxe par excellence. Nous introduisons la syntaxe de dépen-DANCE DE SURFACE et la distinction entre les propriétés fonctionnelles et catégorielles des éléments d'un énoncé. Nous étendons le classement des syntaxèmes à l'ensemble des unités syntaxiques et étudions les CATÉGORIES MICROSYNTAXIQUES et le rôle de la TRANSLATION (Chapitre ??). Le classement des syntagmes nous amène à introduire la notion de RELATION SYNTAXIQUE. Les différentes FONC-TIONS SYNTAXIQUES que peut remplir une unité sont caractérisées et notamment la fonction sujet, qui pose problème, dès qu'on prend en compte les langues ergatives (Chapitre ??). Une attention particulière est portée aux listes ou en-TASSEMENTS PARADIGMATIQUES: nous regroupons sous ce terme différents phénomènes, allant de la coordination à la reformulation, où plusieurs éléments viennent occuper une même position régie (Chapitre ??). L'étude détaillée de l'extraction et du rôle complexe joué par des éléments tels que les pronoms relatifs constitue le dernier chapitre de cette partie (Chapitre ??).

La sixième partie est consacrée à la MACROSYNTAXE, c'est-à-dire l'étude de tout ce qui se situe au-delà de la microsyntaxe, notamment les éléments associés au noyau central de l'énoncé sans pour autant être régis. Nous y définissons l'unité ILLOCUTOIRE qui constitue l'unité minimale du discours et étudions son organisation interne en NOYAU et ADNOYAUX. Nous montrons pourquoi la notion traditionnelle de PHRASE est problématique, notamment parce qu'il y a de la rection au-delà des limites de l'unité illocutoire et qu'à l'inverse des unités non régies peuvent venir s'insérer à l'intérieur d'une unité illocutoire (Chapitre ??).

10 Les termes grammaire, syntaxe et topologie



Le terme grec grammatikē technē (γραμματική τέχνη) désignait « l'art des lettres », lettre dans le sens de « l'écrit »; c'était donc l'étude de l'écrit (et l'étude de sa lecture). Elle s'est ensuite développée en science de l'interprétation des textes, ce que, aujourd'hui, on classe plutôt sous le terme de PHILOLOGIE et que l'on distingue de la GRAMMAIRE. Il y a 3000 ans, bien avant les Grecs, il existait



déjà des études grammaticales (au sens moderne du terme) de langues telles que le sanskrit ou le chinois et au 5^e siècle avant notre ère, le grand linguiste indien Pānini a développé une analyse systématique de la nanosyntaxe du sanskrit. Les Institutiones grammaticae, écrites au 6^e siècle par le grammairien latin Priscien, auront une influence déterminante sur le développement de la grammaire en Europe. Beaucoup de termes grammaticaux encore en utilisation aujourd'hui proviennent de l'étude du latin, lingua franca jusqu'à la fin du moyen-âge.

Le terme syntaxis est également grec : il est composé de syn (συν) 'ensemble' et de táxis (τάξις) 'ordre, arrangement'. La SYNTAXE a désigné l'étude de l'ordre des mots, puis plus largement l'étude de l'organisation des mots dans la phrase. Le terme allemand pour syntaxe est Satzlehre, tout simplement la 'science de la phrase'. Dans la conception traditionnelle, l'analyse syntaxique est limitée d'un côté par le mot, et de l'autre par la phrase : les structures en deçà du mot ne font pas partie de la syntaxe, elles appartiennent à la morphologie; le discours, l'enchaînement de plusieurs phrases, n'est pas le sujet de la syntaxe, mais de la linguistique des textes ou analyse du discours. Notre définition de la syntaxe ne présuppose ni la notion de mot, ni celle de phrase, et considère la MORPHOLOGIE comme l'étude de la combinaison des signifiants des signes (voir la section 4.4 sur Signifié, signifiant, syntactique).

Les termes microsyntaxe et macrosyntaxe ont été forgés en 1990 par les linguistes qui se sont intéressés aux productions orales spontanées (notamment à Aix-en-Provence autour de Claire Blanche-Benveniste et à Fribourg en Suisse autour d'Alain Berrendonner) et ont vu la difficulté que pouvait poser une segmentation en phrases comme à l'écrit. Sur le même modèle, nous proposons le terme nanosyntaxe, à la place du terme morphosyntaxe, pour compléter la partition de l'étude de la syntaxe. Nous introduisons le terme syntaxème pour nommer les unités minimales de la syntaxe, sur le même modèle que les termes morphème et sémantème, désignant respectivement les unités minimales de forme et de sens.

Dans l'usage du terme syntaxe, on est passé de l'étude de l'ordre des mots à, aujourd'hui, l'étude de la combinaison des syntaxèmes et aux structures hiérarchiques qui en résultent. Une des conséquences est qu'il fallait réintroduire un nouveau terme pour l'étude de l'ordre des mots et des syntaxèmes. Nous avons adopté le terme topologie, également d'origine grecque : la TOPOLOGIE est l'étude des topos (τόπος), c'est-à-dire l'étude des lieux. Le terme a été introduit au 19e siècle par les linguistes décrivant l'ordre des mots des langues germaniques à l'aide de gabarits de places : le modèle topologique de l'allemand décrit la façon dont la phrase allemande peut être décomposée en cinq champs, les deuxième et quatrième champs modélisant les positions réservées aux verbes (voir le Chapitre 11 sur *La topologie*).

11 Commentaires sur le plan

Le plan de cet ouvrage amène plusieurs commentaires et permet déjà de se faire une idée des principaux partis pris.

On constatera tout d'abord que nous considérons trois structures « syntaxiques » — la structure topologique, la structure de dépendance de surface et la structure syntaxique profonde — là où la plupart des approches n'en considèrent qu'une : la structure syntagmatique ou analyse en constituants immédiats. Nous aurons plusieurs fois l'occasion de justifier le fait de séparer les informations qui peuvent l'être et donc de dissocier les différents modes d'organisation des différents types d'unités qui apparaissent dans un énoncé.

Autre point : nous donnons une place importante aux interfaces, c'est-à-dire à la correspondance entre les différents niveaux de représentation de l'énoncé. Ceci est principalement dû à notre objectif de modélisation : nous ne souhaitons pas seulement montrer comment un énoncé est structuré, mais aussi comment un locuteur produit ces différentes structures et quels rôles elles jouent dans la production des énoncés. Modéliser la langue, c'est pour nous modéliser comment une personne parle, c'est-à-dire comment elle produit des énoncés dans sa langue en fonction du message qu'elle souhaite communiquer.

La plupart des ouvrages de syntaxe et des cours de linguistique à l'université commencent par l'étude des catégories syntaxiques ou parties du discours, c'est-à-dire la caractérisation de ce qu'est un verbe, un nom, un adjectif, etc. Nous pensons pour notre part qu'une bonne définition des catégories syntaxiques ne peut se faire qu'après avoir dégagé la structure des énoncés et que la caractérisation des catégories repose sur l'analyse distributionnelle des unités syntaxiques à l'intérieur des combinaisons complexes dans lesquelles elles entrent. Autrement dit, on ne peut caractériser les catégories syntaxiques d'une langue par l'étude du simple enchaînement linéaire des unités dans la chaîne parlée : il faut prendre en compte les relations plus complexes qui lient les éléments d'un énoncé et dont l'ordre des mots en surface n'est qu'une projection. La description des liens syntaxiques est au cœur de la syntaxe de dépendance et elle occupe une place centrale dans ce livre. Les catégories syntaxiques seront définies en deux chapitres (4.3 et 5.1) distribués dans les parties consacrées à la nanosyntaxe et à la microsyntaxe.

Cette présentation assez systématique des notions utiles à la syntaxe nous a obligé à définir de nouveaux concepts ou à revoir certains concepts traditionnels.



Q

Selon les cas, nous avons décidé d'utiliser un terme déjà usuel dans un sens un peu différent ou bien nous avons forgé un nouveau terme. Parmi les néologismes que contient cet ouvrage, on notera le terme syntaxème utilisé pour nommer les unités minimales de la syntaxe (qui curieusement ne sont jamais nommées) et nanosyntaxe pour désigner la syntaxe des éléments qui possèdent peu d'autonomie syntaxique. Par ailleurs, d'autres termes peu fréquents dans les manuels de syntaxe occupent ici une position plus centrale, comme sémantème, topologie, macrosyntaxe ou syntaxe profonde.

12 Notions, termes, concepts et définitions



Nous allons illustrer la distinction entre notions, termes, concepts et définitions à partir d'un exemple. Considérons la Définition suivante :

(15) Une phrase est un segment de texte qui se trouve entre deux ponctuations majeures successives.

Un TERME est introduit : phrase. Ce terme est associé à un CONCEPT. Le concept est un objet abstrait, conceptuel, qui n'est accessible qu'à travers la définition qui le caractérise, à savoir « un segment de texte qui se trouve entre deux ponctuations majeures successives ». En associant ce terme et ce concept, nous construisons une notion. Une NOTION est donc un concept nommé ou un terme défini et associé à un concept.

La distinction entre le terme et le concept qui lui correspond est essentielle, puisqu'un même terme peut être associé par différents auteurs à différents concepts. Évidemment, une fois que nous avons associé le terme phrase à un concept, nous pouvons parler de « la notion de phrase », mais il faut être conscient qu'il s'agit d'un raccourci pour désigner « la notion que nous avons nommée phrase et qui ne doit pas être confondue avec une autre notion que d'autres ont pu également nommer phrase et qui peut à l'inverse avoir été nommée autrement par d'autres ».

Il y a plusieurs raisons pour lesquelles, en linguistique, la terminologie n'est pas bien stabilisée et un même terme tend à désigner des concepts divers. Une première raison est que beaucoup de ces termes (mot, phrase, nom, adverbe, sujet, etc.) s'appliquent à des notions qui sont enseignées dès l'école et reçoivent donc des définitions simplifiées et facilement accessibles, qui deviennent inopérante lorsqu'un véritable cadre théorique est développé. Une deuxième raison est qu'il n'est pas possible de définir proprement de telles notions sans se doter d'un

appareil conceptuel complexe et qu'il n'y a pas aujourd'hui de théorie consensuelle sur la nature de la langue et sur les primitives conceptuelles, c'est-à-dire les notions de base à partir desquels des notions plus complexes pourront être définies.

Dans cet ouvrage, nous allons nous attacher à introduire un appareillage théorique rigoureux. Nous introduirons un grand nombre de concepts auxquelles nous associerons bien sûr des termes. Nous avons fait le choix d'utiliser autant que possible les termes courants en linguistique, même lorsque nous décidions de définir une notion un peu différente de la tradition. C'est par exemple le choix que nous avons fait pour le terme syntaxe, auquel nous donnons une acception différente de la tradition. Lorsque nous avons introduit des concepts nouveaux, nous nous sommes permis d'utiliser un terme vacant s'il n'était pas trop éloigné : c'est ce que nous avons fait avec le terme substantif, qui désignait avant ce qu'on appelle aujourd'hui nom et que nous avons attribué à une notion du même ordre, mais différente de celle du nom (voir Chapitre ?? sur Les catégories microsyntaxiques). Lorsqu'aucun terme ne se présentait, nous nous sommes résolus à forger un nouveau terme. Nous avons alors cherché à régulariser la terminologie. C'est ce qui nous a amené à introduire le terme syntaxème à côté des termes morphème et sémantème, ou à introduire le terme nanosyntaxe à côté des termes microsyntaxe et macrosyntaxe.

On peut être en désaccord avec la définition d'une notion. Il est important de voir que ce désaccord peut se situer à trois niveaux bien différents. Revenons sur la notion qui illustre cette section. On peut être en désaccord :

- au niveau proprement définitionnel : on peut considérer que notre définition n'en est pas vraiment une, car elle comprend des termes qui n'ont pas été eux-mêmes définis. Qu'appelle-t-on une ponctuation majeure? Le point-virgule est-il une ponctuation majeure? Qu'entend-on exactement par un segment de texte? S'agit-il juste d'une chaîne de caractères ou bien la phrase est-elle un signe linguistique avec un sens associé au texte proprement dit?
- au niveau théorique ou conceptuel : la notion définie est-elle intéressante d'un point de vue théorique? Certainement pas si on n'élimine pas le cas des points qui suivent une abréviation, comme dans *Georges W. Bush*. Et au-delà, les unités définies par la ponctuation sont-elles bien des unités linguistiques pertinentes?
- au niveau purement terminologique : est-ce bien à ce concept que l'on veut associer le terme *phrase*? Ou inversement, est-ce bien par le terme *phrase* que l'on veut désigner ce concept?

Les problèmes terminologiques sont secondaires : un mauvais choix de terme ne remet pas en cause une théorie. Mais ils peuvent être catastrophiques du point de vue pédagogique et rendre incompréhensible une construction théorique valable par ailleurs. Lorsqu'on introduit un concept, on a principalement deux options terminologiques: utiliser un terme existant, avec le risque que d'autres auteurs l'utilisent avec une acception différente (c'est le cas avec le terme phrase), ou bien forger un nouveau terme, avec le risque d'avoir des termes « barbares » et difficiles à mémoriser. On peut par exemple proposer d'appeler le concept défini plus haut phrase graphique. C'est ce que nous ferons dans la suite de l'ouvrage.

Une phrase graphique est une unité de l'écrit, un segment de texte qui se trouve entre deux ponctuations majeures successives. Les ponctuations majeures sont le point (à l'exception des points utilisés dans les abréviations), le point d'interrogation, le point d'exclamation et les trois petits points. Le point virgule segmente une phrase en sous-phrases.

Les questions théoriques et conceptuelles sont les plus importantes : il est bien sûr crucial d'introduire les bons concepts. C'est l'ensemble des concepts introduits qui définit le cadre théorique à partir duquel un modèle d'une langue particulière pourra être élaboré. La notion de phrase que nous avons introduite est-elle vraiment la plus intéressante du point de vue linguistique? Nous verrons que non, ne serait-ce que parce que l'écrit est une transcription de la langue et qu'une notion de l'ordre de la phrase existe indépendamment de la possibilité d'écrire ou pas (voir la section 1.3 sur les Sons et textes).

Les problèmes définitionnels sont immenses. En général, la définition d'une notion fait appel à d'autres notions. Il faut donc comprendre quels sont les concepts qui doivent être définis avant les autres. Dans notre exemple, nous avons défini la phrase à partir de la ponctuation. Mais comment un locuteur sait-il où mettre des ponctuations majeures quand il écrit? Produit-on des unités de l'ordre de la phrase lorsqu'on parle? Si oui, il existe une unité de l'ordre de la phrase plus fondamentale que celle que nous venons de définir et qui ne se définit pas en fonction de l'écrit et encore moins en fonction de la ponctuation. Si beaucoup d'ouvrages de linguistique commence par la définition de la phrase, nous considérons pour notre part qu'il s'agit d'une notion complexe qui ne peut être définie qu'après avoir introduit la notion de cohésion syntaxique. C'est la raison pour laquelle les notions proches de la phrase seront seulement discutées à la fin de cet ouvrage, dans la sixième et dernière partie. Nous verrons de ce point de vue que derrière la phrase graphique se cache en fait deux types d'unités différentes, ce qui explique qu'il y a différentes façons de ponctuer une même production linguistique.

13 Présentation de l'ouvrage

Cet ouvrage est volontairement découpé en sections n'excédant généralement pas une page. Nous avons essayé de donner à chacune de ces sections autant d'autonomie que possible, de manière à rendre une lecture non linéaire de l'ouvrage aussi facile que possible. Il est néanmoins évident que ce livre a été organisé selon un ordre mûrement réfléchi et que de nombreuses sections ne peuvent être lues sans avoir lu avant les sections qui introduisent certaines notions préalables indispensables.

Certaines sections sont encadrées et présentées dans un style particulier. Ces **encadrés** sont des prolongements du texte principal de différentes natures. Nous avons cinq principaux types d'encadrés :

- des encadrés d'éclairage, prolongeant des points abordés dans le texte principal;
- des encadrés historiques, sur l'origine de certains termes ou de certains concepts;
- des encadrés **techniques**, présentant une élaboration de certaines notions, notamment du côté de la modélisation mathématique;
- des encadrés de **typologie linguistique**, montrant diverses réalisations d'un phénomène ou d'une propriété au travers de la diversité des langues;
- des encadrés sur le **français**, qui reste la langue privilégiée pour illustrer notre propos.

A ces encadrés qui figurent dans le corps du texte s'ajoutent encore quatre autres types d'encadrés placés en fin de chapitre :

- un encadré d'exercices;
- un encadré contenant des éléments de correction de nos exercices;
- un encadré de lectures additionnelles, comprenant en particulier les références citées au cours du chapitre;
- un éventuel encadré de citations originales, lorsque nous avons cité des auteurs qui n'avaient pas écrit en français.

14 Remerciements

Nous remercions les collègues et doctorants qui ont lu des parties du manuscrit et nous ont fait des commentaires qui nous ont parfois amené à réécrire des

parties importantes : Nicolas Mazziotta, Marie-Sophie Pausé, Paola Pietrandrea, Rafaël Poiret.

Nous remercions nos étudiants de la licence de sciences du langage et du master TAL sur qui nous avons testé une grande partie du contenu de cet ouvrage et qui par leurs réactions parfois critiques nous ont permis d'améliorer grandement le texte et d'ajuster le plan de l'ouvrage.

Le contenu de cet ouvrage a fait l'objet de plusieurs articles et communications dans des colloques internationaux, notamment aux conférences bi-annuelles MTT (Meaning-Text Theory), créée en 2003 par Sylvain Kahane et Alexis Nasr, puis Depling (Dependency Linguistics) créée en 2011 par Kim Gerdes, Eva Hajičová et Leo Wanner. Nous remercions les collègues qui nous ont permis de développer nos idées en relisant et critiquant nos articles lors des soumissions ou en nous posant des questions lors des présentations.

Nous remercions également les auteurs qui nous ont précédé et dont la lecture nous a inspiré. Le travail scientifique est cumulatif et la part d'innovation est toujours plus faible qu'on ne le pense. Nous remercions en particulier Claire Blanche-Benveniste, José Deulofeu, Nicolas Mazziotta, Igor Mel'čuk et Alain Polguère avec qui nous avons eu la chance de collaborer et de discuter différents points qui sont développés dans l'ouvrage.



Exercices

Exercice 1 Quelles sont, à côté de la syntaxe, les autres composantes d'un modèle linguistique?

Exercice 2 Les notions de *syntaxe* et grammaire sont souvent confondues. Pouvez-vous donner un élément du modèle d'une langue qui relève de la grammaire mais pas de la syntaxe? Et un élément qui relève de la syntaxe mais pas de la grammaire?

Exercice 3 Qu'appelons-nous la topologie ? L'interface sémantique-syntaxe ? la nanosyntaxe ?

Exercice 4 Pourquoi pensons-nous qu'il est difficile de commencer un ouvrage de syntaxe en définissant les parties du discours?

Exercice 5 De quel type est le premier encadré de cet avant-propos? Que représente le symbole choisi?

Exercice 6 Les mots *syntaxe* et *grammaire* ne sont pas seulement utilisés pour désigner les concepts considérés dans cet ouvrage.

- a) Distinguer les emplois du mot grammaire dans les phrases suivantes :
- Il ne parle pas vraiment le swahili, parce qu'il ne connaît que quelques mots, et il ne connaît pas du tout la grammaire.
- Ca ne se dit pas comme ça, j'ai regardé dans une grammaire.
- Les grammaires de dépendance proposent un formalisme pour représenter les relations qu'entretiennent les mots entre eux.
- C'est incroyable comment un bébé apprend vite la grammaire.
- C'est parce que la grammaire est innée.
- La grammaire française ne permet l'inversion du sujet que dans des cas très restreints.
- On vous enseigne de ne plus faire de fautes de grammaire.
- Les composants fondamentaux des systèmes d'information géographique se conjuguent dans une sorte de grammaire des formes géographiques.
- Le cinéma n'est pas une langue dont il suffirait d'apprendre la grammaire et le vocabulaire.
- La seule grande grammaire italienne disponible en français, celle de Jacqueline Brunet, est à juste titre descriptive (et non normative).
- Après tout, la rigueur de la pensée, on l'apprend avec la grammaire, la philosophie ...
- b) Distinguer les emplois du mot syntaxe dans les phrases suivantes :
- Chaque moteur de recherche a sa propre syntaxe.
- Dans chacun des domaines de la linguistique (syntaxe, phonologie, morphologie et sémantique lexicale), nos connaissances ne sont ni

- apprises ni données.
- Françoise Morvan s'efforce d'inventer une syntaxe française avec parfois des tournures bretonnes.
- J'ai déjà eu l'occasion d'attirer l'attention sur la multiplicité des coquilles, fautes de syntaxe, fautes d'orthographe que l'on trouve dans ce journal.
- Les rédacteurs du Nouvelliste ne maîtrisent qu'approximativement le français, ils se perdent dans la logique de la syntaxe et se noient dans l'abus de vocabulaire.
- J'ai essayé de trouver la bonne syntaxe pour le film.
- Du point de vue de la syntaxe, certaines évolutions sont visibles et prévisibles. Ainsi, le remplacement de "nous" par "on" avec un pluriel. "Mon fils et moi, on est allés au cinéma".
- Envoyez un SMS selon la syntaxe suivante : le mot-clé 'METEO' 'espace' puis le numéro du département dont vous souhaitez connaître la météo.



Lectures additionnelles

Nous recommandons bien sûr la lecture de l'incontournable ouvrage de Lucien Tesnière, *Éléments de syntaxe structurale* et tout particulièrement la première partie. On pourra avant cette lecture lire l'introduction écrite par Sylvain Kahane et Timothy Suborne pour la traduction anglaise de 2015.

En plus de cet ouvrage et de nombreux articles de recherche, quelques livres existent sur la syntaxe de dépendance : Richard Hudson a publié son introduction à la *Word Grammar* en 1984 et plus récemment, en 2006, *Lan*-

guage Networks: The New Word Grammar. En 1986, est paru l'ouvrage The meaning of the sentence in its semantic and pragmatic aspects de Petr Sgall, Eva Hajičová et Jarmila Panevová sur le modèle pragois, qui a conduit au développement de la première banque d'arbre en dépendance, le Prague Dependency Treebank. Il a été suivi par Dependency syntax : Theory and practice d'Igor Mel'čuk, une œuvre fondatrice, mais qui est davantage dédiée à la présentation de la Théorie Sens-Texte, l'une des principales approches théoriques basées sur la syntaxe de dépendance, qu'à la définition de la dépendance. Le récent ouvrage d'Introduction à la linguistique (2014) d'Igor Mel'čuk et Jasmina Milićević, et notamment le second tome consacré à la syntaxe, est à notre connaissance le premier manuel général de linguistique basé sur la dépendance et nous en recommandons la lecture. Notre ouvrage partage en grande partie le point de vue du livre de Mel'čuk et Milićević, bien qu'il ne souhaite pas se placer a priori dans un cadre théorique donné, mais le construire de manière raisonnée. En chinois, Haitao Liu a présenté les grammaires de dépendance en 2009 dans son livre intitulé comme le livre de Mel'čuk1988, Théorie et pratique de la grammaire de dépendance. Dans un cadre formel proche de la grammaire de dépendance, on trouve l'ouvrage de Joan Bresnan sur la Lexical Functional Syntax publié en 2001. On peut encore citer, l'ouvrage de Denis Costaouec et Françoise Guérin de 2007, Syntaxe fonctionnelle : Théorie et exercices, basé sur les travaux d'André Martinet, qui sans se placer réellement dans le cadre de la syntaxe de dépendance, a une approche constructiviste qui se rapproche de la notre.

Joan Bresnan2001 Lexical Functional Syntax, Blackwell, Oxford.

Denis Costaouec & Françoise **Guérin2007** *Syntaxe fonctionnelle : Théorie et exercices*, Presses Universitaires de Rennes.

Richard Hudson1984 Word Grammar, Blackwell, Oxford.

Richard **Hudson2006** *Language Networks : The New Word Grammar*, Oxford University Press.

Sylvain Kahane & Timothy **Osborne2015** Translators' introduction, in Lucien Tesnière, *Elements of structural syntax*, Benjamins, pages xxix–lxxiv (45 pages).

Haitao Liu2009 (Théorie et pratique de la grammaire de dépendance).

Igor **Mel'čuk1988** *Dependency syntax : Theory and* practice, SUNY. Igor Mel'čuk & Jasmina Milićević (**Mel'čukMilićević2014** *Introduction à la linguistique*, 3 volumes, Hermann, Paris.

Petr Sgall, Eva Hajičová & Jarmila **Panevová1986** The meaning of the sentence in its semantic and pragmatic aspects. Springer Science & Business Media.

Lucien **Tesnière1959** Éléments de syntaxe structurale, Klincksieck, Paris.



Corrections des exercices

Corrigé 1 Les principales composantes d'un modèle linguistique sont la sémantique, la syntaxe, la morphologie et la phonologie. On peut ajouter à cela la pragmatique pour l'étude des liens entre le sens linguistique et intentions du locuteur et la phonétique pour l'étude des liens entre les sons de la langue et le signal sonore.

Corrigé 2 La grammaire inclut toutes les composantes du modèle et donc aussi la sémantique et la morphologie. Chacune de ces composantes coupe à travers la grammaire et le lexique. Ce qui relève de la syntaxe sans être de la grammaire est l'étude de la combinatoire des unités lexicales ou de la structure syntaxique interne des locutions. Dans cet ouvrage, nous nous intéressons à la partie grammaticale de la syntaxe (ou, autrement dit, à la partie syntaxique de la grammaire).

Corrigé 3 Toutes ces notions seront étudiées en détail dans cet ouvrage. Mais il peut être bon de fixer les termes dès maintenant. La topologie est

l'étude de l'interface entre syntaxe et ordre linéaire des mots. L'interface sémantique-syntaxe est, comme son nom l'indique l'étude de la correspondance entre sémantique et syntaxe. La nanosyntaxe est la partie de la syntaxe qui s'intéresse aux combinaisons d'unités linguistiques les plus cohésives, celles de l'ordre du mot. Voir les Sections 0.0.8-10.

Corrigé 4 Parce que la définition des parties du discours repose sur une définition préalable de la structure syntaxique (voir section 11).

Corrigé 5 L'encadré ?? est un encadré d'éclairage. Le symbole utilisé est une loupe.

Corrigé 6 Comme tous les termes linguistiques, grammaire et syntaxe peuvent désigner un domaine de la linguistique (4, 5, 7, 11, 13, 15, 16, 18) ou la grammaire ou la syntaxe d'une langue en particulier (1, 6, 14). Par extension, on parle aussi de la grammaire ou syntaxe de systèmes sémiotiques autres que la langue (8,9,12,17,19). Dans grammaire de dépendance (3), le terme désigne un modèle linguistique complet, plutôt que la seule grammaire. On utilise aussi le mot grammaire pour désigner un livre de grammaire (2, 10).

Todo list

unnumbered section	24
check "Saussure" against "de Saussure" in whole book	86
nasalization missing	139
some lists of references are firstname, lastname; others are lastname, first-	
name. Unify.	244
fait footing is discontinous. Maybe choose other verb?	318
check the use of colons and quotations marks in all multiline quotations	
in book	325
Check reference to chapters in book. We currently have 'ch', 'chap', 'cha-	
pitre'; and there are colons and commas found before	334
inflection?	349
number of braces does not match	388
put this in tabular?	408
sg?	423

Première partie Modéliser la langue

Présentation

Cette première partie est consacrée à la modélisation des langues en général. Elle est divisée en trois chapitres. Le premier chapitre essaye de définir la langue, notre objet d'étude, et précise les caractéristiques que nous retenons dans notre modélisation. Le deuxième chapitre montre à travers l'étude de la production d'un énoncé où se situent la syntaxe et la grammaire dans un modèle complet de la langue. Le troisième chapitre caractérise le type de modèles que nous adoptons dans cet ouvrage.

1 La langue : L'objet d'étude de la linguistique

1.1 Parler une langue

Pour comprendre ce qu'est une langue, il faut comprendre à quoi elle sert. Savoir utiliser une langue, c'est être capable de parler dans cette langue et de comprendre ceux qui nous parlent. Parler une langue, c'est être capable de verbaliser n'importe quelle idée, c'est-à-dire transformer un sens en un son dont notre interlocuteur pourra lui-même extraire un sens (voir dans l'encadré qui suit le schéma proposé par Saussure). Si le sens de départ – celui pensé par le locuteur – et le sens d'arrivée – celui construit par le destinataire – sont suffisamment proches, alors, la communication peut être considérée comme réussie.

La langue est donc avant tout un objet qui se trouve dans le cerveau des locuteurs et que l'on peut modéliser par une correspondance entre des sens et des sons. Plus exactement, nous modélisons la langue par une correspondance entre des représentations sémantiques et des textes. Dans la suite, nous utiliserons le terme TEXTE pour désigner les productions langagières qu'elles soient orales, écrites ou gestuelles. Ce terme a l'avantage de ne pas présupposer quelle est la nature du médium utilisé pour communiquer. Il a également l'avantage sur le terme son de renvoyer à une représentation du son et non au son lui-même. Un texte sonore est ainsi la représentation que nous avons du son dans notre cerveau lorsque nous produisons des sons afin de communiquer. L'étude même de la représentation du son — la PHONOLOGIE — ne sera abordée que dans la mesure où elle interfère avec la syntaxe. Nous nous intéressons à la modélisation du mécanisme cognitif (situé dans le cerveau) et nous laissons de côté le mécanisme physique qui permet la production du son à partir du texte (cela concerne la phonétique articulatoire), ainsi que le mécanisme auditif qui permet le décodage du son.

Considérer que la langue est une correspondance, c'est faire abstraction des mécanismes propres à la production ou la compréhension d'un texte. Le locuteur utilise la correspondance du sens vers le texte, tandis que le destinataire effectue le chemin inverse. Nous supposons implicitement que les deux directions, sens

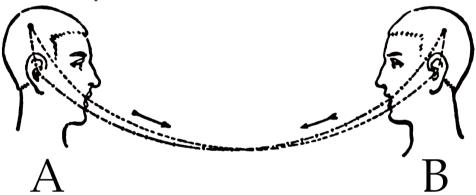
vers texte et texte vers sens, utilisent le même ensemble de connaissances et un mécanisme commun, qui constituent la langue proprement dite.

1.2 La langue comme correspondance sens-texte



On trouve déjà dans le Cours de linguistique générale de Ferdinand de Saussure, publié en 1916, la conception de la langue comme un objet qui fait se correspondre des sens et des « sons ». Voici ce qu'il écrit :

« Pour trouver dans l'ensemble du langage la sphère qui correspond à la langue, il faut se placer devant l'acte individuel qui permet de reconstituer le circuit de la parole. Cet acte suppose au moins deux individus; c'est le minimum exigible pour que le circuit soit complet. Soient donc deux personnes A et B, qui s'entretiennent :



Le point de départ du circuit est dans le cerveau de l'une, par exemple A, où les faits de conscience, que nous appellerons concepts, se trouvent associés aux représentations des signes linguistiques ou images acoustiques servant à leur expression. Supposons qu'un concept donné déclenche dans le cerveau une image acoustique correspondante : c'est un phénomène entièrement psychique, suivi à son tour d'un procès physiologique : le cerveau transmet aux organes de la phonation une impulsion corrélative à l'image; puis les ondes sonores se propagent de la bouche de A à l'oreille de B : procès purement physique. Ensuite, le circuit se prolonge en B dans un ordre inverse : de l'oreille au cerveau, transmission physiologique de l'image acoustique; dans le cerveau association psychique de cette image psychique avec le concept correspondant. Si B parle à son tour ce nouvel acte suivra − de son cerveau à celui de A − exactement la même marche que le premier



et passera par les mêmes phases successives. » (Saussure1916 : 27-28) [c'est nous qui soulignons]

Saussure insiste, quelques lignes plus loin, sur le fait que l'association entre sens et son qui a lieu dans le cerveau ne se fait pas avec le son lui-même, mais avec une représentation de ce son dans le cerveau, que Saussure nomme l'image acoustique:

« [Il faut] distinguer les parties physiques (ondes sonores) des physiologiques (phonation et audition) et psychiques (images verbales et concepts). Il est en effet capital de remarquer que l'image verbale ne se confond pas avec le son lui-même et qu'elle est psychique au même titre que le concept qui lui est associé. » (Saussure1916: 28-29)

Leonard Bloomfield1933, le père de la linguistique américaine, va dans le même sens:

« On produit de nombreuses sortes de bruits vocaux dont on utilise la variété: sous certains types de stimuli, on produit certains sons vocaux et nos compagnons, entendant les mêmes sons, font la réponse appropriée. Pour le dire brièvement, dans la parole humaine, des sons différents ont des sens différents. Étudier la coordination de certains sons avec certains sens, c'est étudier la langue. » [nous soulignons]

Le fait de considérer que l'objet d'étude de la linguistique est de modéliser la correspondance sens-son décrite par Saussure (et appelée par lui association concept-image acoustique) peut être imputé à Žolkovskij et Mel'čuk qui posent en 1967 à Moscou les bases de la Théorie Sens-Texte, dont le nom même est tout à fait explicite sur ce point. Dans son cours au Collège de France en 1997 intitulé Vers une linguistique Sens-Texte, Igor Mel'čuk pose qu'un modèle d'une langue est une correspondance multivoque entre un ensemble de sens et un ensemble de textes, où les textes désignent les images acoustiques des phrases de la langue. Le caractère multivoque est dû au fait qu'un même sens peut être exprimé par différents textes (qui sont alors des « paraphrases » les uns des autres) et qu'un texte peut avoir plusieurs sens (c'est-à-dire être sémantiquement ambigu).

Cette conception de la langue comme une correspondance sens-son est maintenant partagée par la plupart des théories, y compris par la grammaire générative de Noam Chomsky, qui pose, depuis le programme minimaliste élaboré dans les années 1990, qu'un modèle linguistique doit relier sens et sons.

1.3 Sons et textes

Il n'est jamais inutile de rappeler que les langues sont avant tout orales (ou gestuelles comme les langues des signes) et que l'écrit n'est qu'une TRANSCRIP-TION, nécessairement imparfaite et partielle, des productions orales, même s'il tend à avoir son autonomie et à acquérir sa propre codification. D'après Leonard Bloomfield1933, « L'écrit n'est pas la langue, mais simplement une facon d'enregistrer la langue au moyen de marques visibles. Une langue est la même quel que soit le système d'écriture utilisé pour l'enregistrer, exactement comme une personne est la même quelle que soit la façon dont on prend son image. » (Nous traduisons en français toutes les citations. Voir les citations originales en fin de chapitre.) La même idée est reprise par H. A. Gleason1961 : « Une langue écrite est typiquement un reflet, indépendant sous quelques aspects seulement, des langues parlées. En tant qu'image de la parole réelle, elle est inévitablement imparfaite et incomplète. [...] La linguistique doit commencer par une étude approfondie de la langue parlée avant d'étudier la langue écrite. C'est vrai pour des langues avec une longue tradition écrite, comme l'anglais, autant que pour les langues de tribus isolées qui n'ont jamais envisagé la possibilité d'une écriture. » Cent ans avant (dans son ouvrage sur la langue kavi publié en 1836), le philosophe et linguiste Wilhelm von Humboldt écrivait que « la langue, comprise dans son essence réelle, est quelque chose de constant et à la fois, à tout moment, quelque chose de passager. Même sa conservation par l'écriture n'est jamais autre chose qu'un stockage ressemblant à une momie, qui nécessite qu'on cherche à s'imaginer à nouveau le discours vivant. » Cette primauté de l'oralité est constitutive des sciences du langage et la distingue fondamentalement des lettres et de la philologie (l'étude des textes écrits). Les scientifiques n'en ont néanmoins réellement pris conscience qu'au début du vingtième siècle avec les possibilités nouvelles que donnait l'enregistrement des sons et l'intérêt croissant pour les langues sans tradition écrite et notamment les langues amérindiennes; la linguistique s'était jusque-là, à l'exception des travaux des missionnaires qui avaient pour mission d'enseigner la Bible dans des langues inconnues, essentiellement développée par l'étude de langues mortes dont on ne conservait que des traces écrites que l'on souhaitait déchiffrer.

Certaines caractéristiques essentielles des langues sont dues au fait que ce sont des modes de communication oraux. La principale de ces caractéristiques est que la communication orale impose une production linéaire : la CHAÎNE PARLÉE est unidimensionnelle. Les sons doivent être produits les uns à la suite des autres. L'écrit n'imposerait pas cela. La communication écrite fait d'ailleurs un grand

usage de schémas bidimensionnels souvent complexes et la présentation globale d'un texte écrit (titres, paragraphes, encadrés, etc.) joue un rôle non négligeable. Néanmoins, l'écrit traditionnel, et parce qu'il est au départ une transcription de l'oral, a une organisation linéaire, les mots devant se lire à la suite les uns des autres. On notera qu'aujourd'hui, avec l'internet, les textes contiennent une multitude de liens vers d'autres textes et qu'une production textuelle faite de plusieurs pages web n'est plus totalement linéaire : il a une structure en réseau et on parle alors d'HYPERTEXTE. Cette organisation, qui permet différents parcours du texte, existe déjà en partie dans les écrits traditionnels par la présence de notes ou d'encadrés.

Les langues des signes n'ont pas de contraintes de linéarité, puisque les gestes peuvent se développer dans toutes les dimensions de l'espace et différentes parties du corps être utilisées simultanément (mains, position de la tête, regard, orientation du buste, etc.). Il y a néanmoins une contrainte temporelle dans la communication qui veut qu'il y ait une certaine séquentialité des signes.

Le caractère linéaire de la chaîne parlée est lui-même en partie contourné par l'utilisation de la prosodie : les locuteurs ajoutent à la suite des sons distinctifs élémentaires qu'ils produisent — les phonèmes — de l'information en modulant leur mélodie et en jouant sur l'intensité du signal et la durée des phonèmes. Ceci permet non seulement de communiquer certaines émotions, mais aussi de structurer la chaîne parlée en faisant apparaître des regroupements ou des changements de plan. On retrouve à l'écrit une transcription de la prosodie par la PONCTUATION (point, virgule, :,?,! On peut imaginer qu'un langage purement écrit aurait développé un tout autre système et on voit que la ponctuation n'est rien d'autre qu'un marquage partiel et imparfait de la prosodie. On notera que, avec le récent développement du dialogue par écrit, le système de ponctuation s'est enrichi des smileys, émojis et autres émoticônes (o, o et plein d'autres).

Nous aurons plusieurs fois l'occasion dans cet ouvrage de montrer que les faits de langue se comprennent bien mieux lorsqu'on se place du côté de l'oral et que notre objet d'étude est d'abord la langue parlée, même si, par commodité, dans un ouvrage écrit, il est souvent plus facile de donner des exemples écrits.

1.4 Langue et variations 🕥



On observe de nombreuses variations entre locuteurs d'une même langue : accent, choix lexicaux, constructions grammaticales, etc. Ces variations sont dues à de multiples facteurs : le degré d'apprentissage de la langue, l'époque à laquelle vit le locuteur, le lieu où il vit, le contexte social dans lequel il s'exprime (la famille, le travail, la radio ...), etc. Ainsi l'allemand est pris entre deux langues très proches, le néerlandais et le suisse allemand et on observe un continuum de variétés d'allemand lorsqu'on se rapproche de ces deux zones linguistiques. Pour le français, on observe également des différences de parler suivant les régions, entre le nord et le sud de la France (le système phonologique, notamment, est différent), mais aussi entre la France, la Belgique, la Suisse, le Québec ou les pays francophones d'Afrique. Et à chaque fois, on observe un continuum de parler entre divers extrêmes. Il en va de même pour l'évolution des langues : la limite entre le latin et les langues romanes d'aujourd'hui (italien, espagnol, catalan, francais, roumain, corse, etc.) n'est pas une frontière tranchée : le latin a évolué de génération en génération jusqu'à perdre ses désinences casuelles (qu'on ne retrouve dans aucune des langues romanes), puis divers groupes de locuteurs du haut-latin se sont retrouvés isolés au Moyen-Age et ont développés les dialectes que sont nos langues d'aujourd'hui. Le français, ancienne langue d'oïl, a subi l'influence de locuteurs d'origine germanique et la syntaxe de l'ancien français est beaucoup plus proche de celle des langues germaniques d'aujourd'hui (allemand, néerlandais, scandinave) que du latin classique. La situation est identique partout. Si l'on appelle chinois aussi bien le chinois classique que le mandarin actuel, ces langues n'en sont pas moins éloignés que le latin et l'italien. Il existe d'ailleurs aujourd'hui en Chine au moins quatre langues différentes, qui bien que partageant la même écriture, sont plus différentes entre elles que ne le sont les langues romanes entre elles.

D'une certaine façon, ces variations ne nous intéressent pas : nous décrivons un système unique, celui d'un locuteur donné à un moment donné ou au moins la langue d'un groupe de locuteurs qui communiquent usuellement entre eux. Ce qui nous intéresse avant tout, c'est la cohérence interne de ce système. Néanmoins, les variations possibles de ce système peuvent parfois nous intéresser : elles nous permettent en particulier de comprendre certaines « bizarreries » du système qu'on ne saurait expliquer sans prendre en compte le fait qu'il s'agit d'un système en évolution. Par exemple, la non-correspondance entre les unités de forme et les unités de sens (question que nous développerons en long et en large dans la partie 2 consacrée aux Unités de la langue) ne peut s'expliquer sans prendre en compte comment de nouveaux termes sont créés par les locuteurs et comment d'autres disparaissent. Il faut d'ailleurs distinguer deux types de changements dans les langues : les changements lexicaux et les changements grammaticaux. Tout locuteur modifie constamment son vocabulaire, acquérant de nouvelles unités lexicales et cessant d'en utiliser certaines. Par contre, il est peu probable qu'une fois l'enfance passée et l'acquisition complète de la grammaire effectuée des changements se produisent dans le système grammatical. Il est plus raisonnable de penser que les changements grammaticaux ont lieu lors du passage de relais d'une génération à l'autre : pour une raison ou une autre, l'apprenant va faire une analyse différente des productions langagières de ses « instructeurs » et construire une grammaire différente de la leur. Le cas le plus radical est celui d'un groupe d'apprenants d'une autre langue maternelle qui va projeter sur la langue qu'il apprend des constructions de sa langue maternelle. Les langues romanes et tout particulièrement l'ancien français sont ainsi des formes du latin dont une partie de l'évolution est due à l'assimilation d'apprenants de langue germanique.

1.5 Sens et intention communicative

Notre définition de la langue n'est compréhensible que si on s'entend quelque peu sur ce qu'est le sens. Pour cela nous allons partir de la définition que donne Leonard Bloomfield dans *Language*, son ouvrage fondateur de 1933.

Pour Bloomfield, il y a acte de langage lorsque Jill a faim, qu'elle voit une pomme dans un arbre et qu'au lieu de grimper dans l'arbre la cueillir, elle produit un son avec son larynx, sa langue et ses lèvres et que c'est Jack qui cueille la pomme et la lui apporte. Autrement dit, face à un stimulus S (Jill a faim et voit une pomme), il y a deux voies pour arriver à la réaction R :

- la voie directe S R, où Jill grimpe dans l'arbre,
- et la voie indirecte S r s R, où une réaction linguistique r se substitue à la réaction mécanique de Jill et où le stimulus s qu'elle provoque donne la réaction R chez Jack.

Reprenons la citation de Bloomfield, donnée dans l'encadré ?? sur *La langue comme correspondance sens-texte*, sous ce nouvel éclairage :

« On produit de nombreuses sortes de bruit vocaux dont on utilise la variété: sous certains types de stimuli, on produit certains sons vocaux et nos compagnons, entendant les mêmes sons, font la réponse appropriée. Pour le dire brièvement, dans la parole humaine, des sons différents ont des sens différents. Étudier la coordination de certains sons avec certains sens, c'est étudier la langue. » (Bloomfield1933 : 27)

Jusque-là, nous sommes parfaitement d'accord. Reste à définir le sens :

« En produisant une forme linguistique, un locuteur incite son interlocuteur à répondre à une situation; cette situation et la réponse qu'elle déclenche

sont le *sens linguistique* de la forme. Nous supposons que chaque forme linguistique a un sens constant et défini, différent du sens de n'importe quelle autre forme linguistique de la même langue. » (**Bloomfield1933** : 165)

Là, nous ne sommes plus en accord. Nous pensons qu'il faut absolument séparer la situation du sens linguistique. Dans la même situation, Jill peut produire des énoncés de sens différents comme « *Pourrais-tu me cueillir cette pomme*? » ou « *Apporte-moi cette pomme!* » ou des énoncés moins coercitifs comme « *J'ai faim.* » ou « *Regarde cette belle pomme!* », qui peuvent néanmoins amener la même réaction de Jack. Inversement, dans une tout autre situation, par exemple face à une nature morte dans un musée, Marie peut dire à Pierre « *Regarde cette belle pomme!* » et cet énoncé a, pour notre définition du sens, le même sens que l'énoncé « *Regarde cette belle pomme!* » de Jill, qui procédait pourtant d'intentions complètement différentes.

Autrement dit, nous distinguons clairement trois objets :

- le contexte d'énonciation, c'est-à-dire les caractéristiques extérieures de la situation où est produit l'énoncé : qui parle à qui? où et quand? dans quelles circonstances? etc.
- les intentions communicatives du locuteur, c'est-à-dire les buts que se fixe le locuteur, les informations qu'il souhaite communiquer sur tel ou tel objet dans le contexte, etc.
- le sens linguistique de l'énoncé, c'est-à-dire, indépendamment du contexte et des intentions du locuteur, le contenu de son message, les éléments de sens qu'il a choisi pour communiquer l'information, désigner tel objet, etc.
 Le sens linguistique, tel que nous l'envisageons, est très proche du texte, puisqu'il contient déjà les sens des différentes unités du texte.

La phase d'élaboration du contenu d'un message, c'est-à-dire le passage d'intentions communicatives dans un contexte donné à un sens linguistique s'appelle la PLANIFICATION du message. On considère généralement que l'étude de la planification ne relève pas de la linguistique et que cette étape reste relativement indépendante du langage, dans la mesure où toutes les langues permettent d'exprimer à peu près tous les sens (comme le montre l'absence d'obstacles majeurs à la traduction d'une langue à l'autre, sauf lorsqu'il s'agit de concept absent d'une culture à l'autre). Laurence Danlos, l'une des pionnières de la génération automatique de textes, propose d'appeler la planification le *Quoi dire*, qu'elle oppose au *Comment le dire*, qui constitue la langue proprement dite.

Nous savons bien sûr que la planification est en partie guidée par le stock lexical que chaque langue propose et nous allons voir un peu plus loin (section 1.10) que la planification joue quand même un rôle dans l'organisation des énoncés et



la syntaxe, même si dans cet ouvrage nous étudierons essentiellement le passage du sens au texte. L'étude même des sens linguistiques — la sémantique — ne sera abordée que dans la mesure où elle interfère avec la syntaxe.

1.6 Mots et pensée

La question se pose de savoir si on peut penser sans mots, c'est-à-dire si on peut manipuler des concepts sans les verbaliser, même dans sa tête. Nous pensons que oui. Le bricoleur qui répare son moteur pense à toutes les pièces du moteur qu'il manipule sans avoir nécessairement une idée de comment les appeler et il effectue un grand nombre d'actions bien réfléchies qu'il aurait bien du mal à expliquer en mots. Plus on va vers une pensée abstraite, plus on pourrait penser que les sens doivent se confondre avec les mots. Pourtant, le mathématicien qui fait une démonstration n'a pas toujours les mots pour exprimer les concepts qu'il manipule et une part de son activité est justement d'isoler ces concepts et de leur donner des noms : ensemble, fonction, continuité, etc. Ici il est clair que le concept précède dans la pensée le terme qui lui correspond. Nous faisons l'hypothèse qu'il existe une forme abstraite de pensée sans mots et que la langue est l'ensemble des connaissances qui nous permet d'exprimer cette pensée. Cette hypothèse reste aujourd'hui plus ou moins invérifiable et est controversée. Pour une discussion très lisible de ces idées, voir l'ouvrage de Steven Pinker, Comment fonctionne l'esprit, Éditions Odile Jacob, 2000.

1.7 Sens lexicaux et traduction 🕏

Les sens exprimables par des mots simples varient d'une langue à l'autre, ce qui élimine tout espoir de traduction exacte. Par exemple, l'anglais possède une unité lexicale MELON qui dénote aussi bien le melon que la pastèque et le francais ne possède pas d'équivalent. Le verbe ESPERAR de l'espagnol couvre à la fois les sens 'espérer' et 'attendre' du français et il existe un continuum entre les deux sens : 'esperar' ne sera donc pas ambigu mais « sous-spécifié » en ce qui concerne le degré de joie et de certitude qu'a la personne qui « espera ». À l'inverse, le verbe AIMER couvre ce que l'anglais exprime avec LOVE et LIKE, car encore une fois, les locuteurs du français considère un continuum entre ces sens. Il existe aussi des sens lexicaux propres à une langue : par exemple, le nom allemand schadenfreunde désigne la joie que procure le dépit mérité des autres et n'a pas d'équivalent en français, bien que le concept puisse être universellement compréhensible. De la même façon, un verbe français comme s'ACCOUDER n'aura pas de meilleur équivalent dans beaucoup d'autres langues (par exemple en anglais ou en allemand) qu'une traduction littérale de 's'appuyer sur les coudes'.

1.8 Langue, linguistique et modélisation

Nous allons définir un certain nombre de termes dont nous aurons besoin et que nous avons déjà commencé à utiliser.

Un locuteur ou sujet parlant est quelqu'un qui parle, c'est-à-dire quelqu'un qui cherche à communiquer avec d'autres gens en produisant des paroles ou un texte écrit. Il s'adresse à des destinataires ou interlocuteurs.

Une langue est un système de signes conventionnels partagés par un certain nombre de personnes et qui leur permet de communiquer entre elles. Ces signes s'assemblent pour former des mots, des phrases et des discours. Une langue est à la fois un **objet individuel** — c'est l'ensemble des connaissances stockées dans notre cerveau qui nous permettent de parler (dans cette langue) — et un **objet collectif** et **social**, puisque ces connaissances sont partagées par un certain nombre de personnes, qui sont les locuteurs de cette langue.

La FACULTÉ LANGAGIÈRE est l'aptitude que nous avons à apprendre et utiliser les langues. Le cerveau est l'organe de la faculté langagière (avec le système phonatoire que le cerveau commande pour produire des sons). Lorsqu'on parle de la langue, et non plus d'une langue particulière, on



fait généralement référence à la faculté langagière et ce qu'elle va imposer comme traits communs à l'ensemble des langues possibles.

La LINGUISTIQUE est la science qui étudie les langues du monde. Comme beaucoup de scientifiques, nous considérons que la linguistique inclut l'étude de la faculté langagière, c'est-à-dire l'étude de la production langagière et de l'apprentissage d'une langue par l'enfant. La linguistique est alors quasiment une branche de la psychologie.

La linguistique produit des MODÈLES DES LANGUES et de la faculté langagière. Ces modèles doivent être capables de simuler un ACTE DE LANGAGE, c'est-à-dire la façon dont un locuteur produit un texte à partir d'un sens qu'il veut exprimer et la façon dont son interlocuteur reconstruit un sens à partir de ce texte.

Tout modèle se situe dans un CADRE THÉORIQUE. Décider que la langue est une correspondance est un choix théorique. Décider ce qui est mis en correspondance par la langue, c'est-à-dire ce que sont la représentation sémantique et le texte relève aussi de choix théoriques. C'est le cadre théorique qui caractérise notamment l'objet d'étude. Pour reprendre une formule de Saussure1916 devenue fameuse, « Bien loin que l'objet précède le point de vue, on dirait que c'est le point de vue qui crée l'objet. ».

1.9 Langue et parole, compétence et performance



Saussure oppose deux notions fondamentales qu'il nomme langue et parole :

« Entre tous les individus ainsi reliés par le langage, il s'établira une sorte de moyenne: tous reproduiront, — non exactement, mais approximativement — les mêmes signes unis aux mêmes concepts. [...]

La langue n'est pas une fonction du sujet parlant, elle est le produit que l'individu enregistre passivement. [...] Elle est la partie sociale du langage, extérieure à l'individu, qui à lui seul ne peut ni la créer ni la modifier; elle n'existe qu'en vertu d'une sorte de contrat passé entre les membres de la communauté. [...]

La **parole** est au contraire un acte individuel de volonté et d'intelligence, dans lequel il convient de distinguer :

1° les combinaisons par lesquelles le sujet parlant utilise le code de la langue en vue d'exprimer sa pensée personnelle;

2° le mécanisme psycho-physique qui lui permet d'extérioriser ces combinaisons. »

(Saussure1916: 29)

La parole, au sens de Saussure, couvre deux notions qu'il convient de séparer. Nous préférons parler de productions langagières pour la première notion, c'est-à-dire les énoncés réellement produits par des sujets parlants, tandis qu'on préférera appeler la deuxième notion la faculté langagière. Alors que la parole est un objet individuel, la langue est un objet social par excellence, mais c'est aussi la trace qu'a imprimée cet objet dans le cerveau de chacun de nous, objet collectif, donc, qui n'existe que par la somme de ses traces individuelles.

L'opposition entre compétence et performance proposée par Chomsky1965 nous renvoie à l'opposition entre langue et parole, mais il convient de les distinguer : la compétence ne se confond pas avec la langue, ni la performance avec la parole. La compétence désigne notre compétence passive à savoir utiliser la langue, mais aussi à l'acquérir. Elle se divise en une compétence innée, qui peut se confondre avec la faculté langagière, et une compétence acquise, qui peut se confondre avec la langue en tant que trace individuelle d'une langue dans notre cerveau.

La performance est l'usage proprement dit de la langue. La parole est le produit de la compétence et de la performance : nous avons la compétence de produire des énoncés supposément parfaits, mais divers facteurs (notre état émotionnel, des éléments qui vont nous distraire, la recherche d'un message approprié, etc.) vont faire que notre énoncé ne sera pas aussi parfait qu'il aurait pu l'être. Ceux qui veulent décrire la langue, comme nous, vont essayer de séparer ce qui relève d'un manque de compétence de ce qui relève d'une erreur de performance. La chose est loin d'être évidente, notamment lorsqu'on touche aux limitations mémorielles. Par exemple, Chomsky a beaucoup insisté sur le caractère récursif de la langue : une proposition peut contenir une proposition subordonnée (La personne [que le chien a mordu] est à l'hôpital) et un tel enchâssement peut être itéré (La personne [que le chien [auquel le garçon a donné un os] a mordu] est à l'hôpital). Mais cette dernière phrase est difficilement compréhensible et une insertion de plus dépasse nos capacités d'analyse en situation de

communication ordinaire (*La personne que le chien auquel le garçon qui habite au coin de la rue a donné un os a mordu est à l'hôpital*). Défaut de compétence ou de performance? (Voir Exercice 3.)

1.10 La planification

Nous voudrions montrer ici que la planification (voir définition dans la section 1.5 sur Sens et intention communicative) peut avoir des incidences non négligeables sur la nature du texte produit et que dans l'absolu il faut l'inclure dans notre objet d'étude. En particulier, on ne peut pas en général considérer que le locuteur construit le contenu de son message avant de le transformer en un énoncé, c'est-à-dire qu'il planifie complètement avant de produire un énoncé. Les choses ne se passent généralement pas comme ça et le locuteur élabore le contenu de son message au fur et à mesure de l'énonciation. Les productions orales présentent en particulier de nombreux indices de la planification en cours. (C'est moins net à l'écrit, puisque le scripteur à la possibilité de revenir en arrière pour corriger sa production.) Comme le dit Claire Blanche-Benveniste 1990, « Lorsque nous produisons des discours non préparés, nous les composons au fur et à mesure de leur production, en laissant des traces de cette production. [...] L'étude de ces traces est en elle-même un sujet d'observations; on y voit la production de langage en train de se faire. [...] Une observation attentive permet de voir comment nous procédons, quelles unités nous utilisons pour faire avancer nos discours, quelles tenues en mémoire nous avons, à la fois pour les morceaux déjà énoncés et pour ceux que nous projetons d'énoncer. On peut ainsi observer comment se fait la mise au point des syntagmes, la recherche des « bonnes dénominations », et le travail constant d'évaluation que nous faisons sur nos propres discours. »

Nous allons montrer trois phénomènes qui illustrent l'influence de la planification sur la structure de l'énoncé. Les exemples qui suivent sont des retranscriptions fidèles de productions orales attestées; dans ces transcriptions figurent absolument tous les mots prononcés par le locuteur, y compris les bribes et répétitions dues aux hésitations du locuteur.

Les premiers indices de la planification en cours sont les nombreuses amorces que le locuteur fait et auxquelles il renonce momentanément ou définitivement. Dans l'exemple suivant, la locutrice semble ne pas trouver tout de suite la bonne formulation; elle hésite et répété *les* pour se donner du temps, amorce la production de *les capitales*, mais s'y prend quand même à deux fois pour finalement proposer une reformulation par *les grandes villes*:

À chaque fois, on obtient une BRIBE, c'est-à-dire un segment inachevé qui est

et je voulais pas aller à Addis Abeba puisque **les les les c- les capitales les grandes villes** ne me disaient rien du tout

ensuite corrigé par le locuteur. La disposition suivante du texte, dite ANALYSE EN GRILLE, permet de mettre en évidence l'entassement (voir le Chapitre ??) dans la même position syntaxique de la bribe et du segment qui vient la remplacer :

(1) et je voulais pas aller à AA puisque les

les

les

les c-

les capitales

les grandes villes ne me disaient rien du tout

Plus étonnant est le phénomène de la GREFFE bien étudié par José Deulofeu (voir la partie 6). L'exemple suivant en fourni deux :

(2) on avait critiqué le le journal de je crois que c'était le Provençal on l'avait critiqué par rapport à ou le Méridional par rapport à la mort de comment il s'appelle ... pas Coluche l'autre

À deux reprises, ici, le locuteur ne trouve pas le nom qu'il cherche et il vient greffer un énoncé qui pourrait fonctionner comme un énoncé autonome et qui est ici inséré dans l'énoncé principal dont il assure la complétion. Nous reprenons le texte précédent en le disposant selon les principes de l'analyse en grille et en indiquant les greffes en gras.

(3) on avait critiqué le le journal de je crois que c'était le Provençal on l'avait critiqué par rapport à ou le Méridional par rapport à la mort de comment il s'appelle pas Coluche l'autre

On notera que la première greffe s'entrelace avec l'énoncé principal, ce qui semble montrer que le locuteur poursuit en parallèle la planification de son énoncé principal et de la greffe. On voit à nouveau ici un fonctionnement par entassement de segments similaires ou identiques (on avait critiqué le journal - on l'avait critiqué, par rapport à - par rapport à).

Un dernier phénomène illustre bien le fait que la planification a lieu en même temps que l'énonciation : les parenthèses. On appelle parenthèse tout énoncé qui vient s'insérer dans l'énoncé principal et qui est marqué par un changement de registre (une modification de l'intonation) qui le détache nettement de l'énoncé principal. Dans le texte suivant, nous avons indiqué les parenthèses entre parenthèses et nous avons directement disposé le texte en grille :

(4) donc pour essayer un petit peu de sortir cette personne de la misère (car c'est vraiment un petit peu semblable aux Misérables de Victor-Hugo)
nous essayons tant bien que mal de lui faire comprendre que sa cabane dans quelques années (entre parenthèses, elle a 79 ans)
quand elle aura des difficultés (ce qu'on espère pas)
des difficultés à se déplacer ou à évoluer (c'est-à-dire qu'il y a énormément d'escaliers à monter pour arriver à sa cabane)
donc le jour où elle ne pourra plus se déplacer
ou qu'elle sera malade un petit peu plus sévèrement,
on essaye de lui faire comprendre qu'elle ne pourra plus vivre dans cette cabane

On peut résumer le contenu de ce texte ainsi : comme cette personne a 79 ans et qu'il y a énormément d'escaliers pour arriver à sa misérable cabane, il faut lui faire comprendre qu'elle ne pourra plus y vivre dans quelques années. On voit que deux informations essentielles ('elle a 79 ans' et 'il y a énormément d'escaliers') ont été ajoutées à la volée, ce qui a finalement obligé le locuteur à abandonner sa première proposition (sa cabane, en gras dans le texte, est le sujet d'un verbe originalement planifié qui ne vient jamais) et à reprendre par une proposition équivalente (nous essayons tant bien que mal de lui faire comprendre que sa cabane \rightarrow on essaye de lui faire comprendre qu'elle ne pourra plus vivre dans cette cabane). Il est probable qu'à la première lecture (qui aurait normalement dû être une écoute) de ce texte, vous ne vous êtes pas rendu compte que la première proposition avait été laissée inachevée : ceci montre le caractère très naturel de telles constructions et le fait que le destinataire est habitué à « corriger » les « erreurs » dues à la planification.

Les linguistes considèrent généralement que la planification est hors de leur objet d'étude et que la langue constitue uniquement le passage du contenu du message, c'est-à-dire le sens, à un énoncé. Les exemples précédents montrent que la planification, ou plutôt les problèmes de planification, laisse de nombreuses

traces en surface et qu'il est donc difficile d'en faire abstraction, surtout si on étudie les productions orales. À l'écrit, par contre, les défauts de planification sont gommés par les passages successifs du rédacteur et la possibilité d'interrompre la rédaction pendant la planification. Ceci est encore une raison de préférer l'étude de l'oral à celle de l'écrit, car on trouve à l'oral davantage d'indices de la façon dont les locuteurs « travaillent », alors les corrections successives sont invisibles à l'écrit.

1.11 Corpus et introspection

Il existe deux moyens d'obtenir des faits de langue pour le linguiste. Le premier est de collecter des textes déjà produits. Un ensemble de textes est appelé un corpus. Le plus grand corpus disponible est le web et les moteurs de recherche constituent un assez bon moyen de récolter les données que l'on cherche, même s'il faut savoir trier ces données selon le type de page (site d'information, blog, forum, chat, etc.) et le type d'auteur (locuteur natif, génération automatique, traduction automatique, etc.). Il existe des corpus plus spécialisés, comme les bibliothèques numériques d'ouvrages classiques, les archives des grands journaux, les encyclopédies en ligne ou les revues scientifiques. Les linguistes constituent des corpus pour leur besoin, notamment des corpus de productions orales dont les textes sont minutieusement retranscrits à l'écrit (nous en avons donné des exemples dans la section précédente). Certains de ces corpus concernent des populations particulières : enfants en phase d'acquisition, apprenants d'une seconde langue, aphasiques, etc. Tous ces paramètres constituent le GENRE de la production textuelle. Un bon corpus doit comporter ce type d'informations, qu'on appelle les MÉTADONNÉES, c'est-à-dire les données qui concernent le corpus et se trouvent à côté des données proprement dites. En plus des métadonnées, certains corpus sont agrémentés d'annotations diverses permettant une meilleure étude des structures des énoncés. On trouve également des corpus alignés de différentes langues (qu'on appelle corpus multilingues ou bitextes) très utiles pour développer des modèles pour la traduction.

Le deuxième moyen d'étude est l'introspection. Il s'agit de construire artificiellement des énoncés et d'en faire juger l'ACCEPTABILITÉ par des locuteurs natifs. Ce moyen permet de tester toutes les variantes imaginables d'un phénomène et surtout de vérifier les limites d'un phénomène en produisant des énoncés jugés inacceptables. Une autre raison qui peut justifier le recours à l'introspection est que, sur corpus, on rencontre beaucoup d'erreurs de performance, qui font que certains énoncés seraient jugés inacceptables même par ceux qui les ont produits.

Il est donc nécessaire de garder un esprit critique et de savoir filtrer les résultats.

Un énoncé produit dans des conditions normales de production est dit ATTESTÉ, par opposition à un énoncé construit par le linguiste. Même si nous ne rejetons pas l'introspection et l'appel au jugement des locuteurs, nous considérons que l'étude des corpus restent le meilleur moyen d'accéder aux données et d'éviter de passer à côté de phénomènes importants.

1.12 Acceptabilité

Parmi toutes les phrases bizarres qu'un linguiste rencontre ou construit, il est souvent difficile de classer les phrases en bonnes et mauvaises. On constate plutôt une gradation de « qualité » qu'un jugement binaire en bon et mauvais.

Considérons les énoncés construits suivants :

- (5) a. C'est un film que je sais que tu n'hésiteras pas une seconde à regarder.
 - b. C'est un film que je me demande quand tu regarderas.
 - c. C'est un film que je ne sais pas si tu accepteras que je regarde.
 - d. C'est un film que je me demande jusqu'où tu es prêt à regarder.
 - e. C'est un film que je dormais quand tu regardais.

On ressent facilement que la phrase (A) est meilleure que la phrase (B), qui elle-même est meilleure que (C). On constate que la phrase (E) est clairement inacceptable, tandis qu'on peut se demander si (D) l'est ou pas. Il est d'usage de noter l'acceptabilité des énoncés par des symboles allant de l'absence de symbole signifiant l'acceptabilité au symbole * signifiant l'inacceptabilité, en passant par les symboles ? (léger doute), ?? (doute sérieux) et ?* (inacceptabilité probable). Pour nos exemples, nous aurions donc :

- (6) a. C'est un film que je sais que tu n'hésiteras pas une seconde à <mark>r</mark>egarder.
 - b. [?]C'est un film que je me demande quand tu regarderas.
 - c. ??C'est un film que je ne sais pas si tu accepteras que je regarde.
 - d. ?*C'est un film que je me demande jusqu'où tu es prêt à regarder.
 - e. *C'est un film que je dormais quand tu regardais.

Les marques d'acceptabilité ne sont pas absolues et doivent plutôt être interprétées comme relatives (c'est-à-dire qu'un énoncé marqué?? est plus acceptable qu'un énoncé marqué?*).

Acceptabilité et grammaticalité

unnumbere

1 La langue : L'objet d'étude de la linguistique

Lorsqu'on étudie la syntaxe, il est important de faire abstraction des problèmes qui viennent de la sémantique. Comparons les deux énoncés suivants :

- (7) a. D'élégants chevaux blancs courent librement.
 - b. D'incolores idées vertes dorment furieusement.

Évidemment, l'énoncé (8), traduit d'un célèbre exemple de Noam Chomsky (Colorless green ideas sleep furiously), est plus que bizarre et il est difficile de trouver un contexte, autre que poétique, où cet énoncé aurait un sens approprié. Pourtant d'un strict point de vue syntaxique, l'énoncé (8) est identique à (7) et peut être jugé comme tout à fait GRAMMATICAL.

Voici un autre exemple (attesté) d'une phrase grammaticale, mais incompréhensible, car trop complexe et appartenant à un langage de spécialité dont le vocabulaire nous est inhabituel :

(8) À mon avis, à l'exception de l'effet des éventuels redressements que j'aurais pu juger nécessaires si j'avais pu m'assurer de l'intégralité des produits dont il est question au paragraphe précédent, ces états financiers présentent fidèlement, à tous égards importants, la situation financière de la société au 31 décembre 1996 ainsi que les résultats de son exploitation et l'évolution de sa situation financière pour la période de douze mois terminée à cette date selon les principes comptables généralement reconnus.

Lorsqu'un énoncé est grammatical, mais jugé INAPPROPRIÉ, nous utilisons le symbole #. Ainsi en réponse à la question « *Qui a mangé les framboises*? », la réponse suivante sera jugée inappropriée :

(9) [#]C'est les framboises que Pierre a mangées.

Autre exemple : quand on parle des expressions figées, on peut relever que MBRISER LA GLACEM (au sens de 'dissiper la gêne') se passive, mais pas MPERDRE LES PÉDALESM (au sens de 'perdre le contrôle de soi-même') :

- (10) Marie a brisé la glace.
- (11) La glace a été brisée par Marie
- (12) Marie a perdu les pédales.

(13) # Les pédales ont été perdues par Marie.

Cette dernière phrase est grammaticale, mais elle a perdu le sens figé (la seule lecture possible est littérale : on parle vraiment de pédales et d'une perte) et elle n'a donc plus le sens approprié.

Enfin, lorsqu'une combinaison morphologique est jugée inappropriée, car absente du stock lexical actuel, nous utilisons le symbole °: °tournement, °bravitude, °aspire-poussière, °loup-chien, etc.

1.13 Parler et comprendre

Si les textes (oraux ou écrits) sont le principal moyen d'accès à la langue, il ne faut pas oublier que la description des textes n'est pas notre finalité. Un texte est une production langagière et c'est bien la production du texte qui, derrière le texte, nous intéresse.

Une langue est une correspondance entre sens et textes et donc modéliser la langue ce n'est pas seulement modéliser les textes de cette langue, mais modéliser la correspondance entre sens et textes. Or il y a deux façons d'effectuer cette correspondance : soit on passe du sens au texte, c'est-à-dire qu'on parle, soit on passe du texte au sens et l'on est dans une situation d'analyse et de décodage du texte.

La plupart des études partent des textes et donc modélisent la langue dans le sens de l'analyse. Nous pensons, à la suite d'Igor Mel'čuk, qu'il est préférable d'étudier la production et de travailler dans le sens de la synthèse (on parle encore de génération de textes quand la production est automatisée). Ceci peut paraître parfois délicat, car pour étudier la production, il faut partir d'une intention communicative et donc commencer par construire un sens. Mais c'est le seul moyen de comprendre quelles sont les contraintes qui s'exercent sur le locuteur et modèlent la langue. Nous allons illustrer notre propos en étudiant un exemple de production dans le chapitre suivant.



Exercices

Exercice 1 S'intéresse-t-on aux variations individuelles entre locuteurs d'une même langue?

Exercice 2 Le symbole ?* veut-il dire qu'on ne croit pas que la phrase est agrammaticale ?

Exercice 3 Nous avons terminé l'encadré ?? en nous demandant si le fait que la phrase *La personne que le chien auquel le garçon qui habite au coin de la rue a donné un os a mordu est à l'hôpital* était incompréhensible mettait en évidence un problème de compétence ou de performance. Quel élément de réponse peut-on apporter?

Exercice 4 Quel phénomène intéressant observe-t-on dans les énoncés suivants du point de vue de la planification?

mais voilà ça c'est une sorte de mélange très très très étrange et qui moi me je sais pas pourquoi me renverse littéralement

vous allez passer devant la poste qui sera à votre droite et un peu plus loin euh je sais pas à quel niveau c'est exactement Habitat et euh la Chambre de de Commerce à votre gauche

Exercice 5 La société humaine repose sur la croyance. Cette idée est très bien développée dans *Sapiens - Une brève histoire de l'humanité* de Yuval Noah Harari. La question dépasse très largement la croyance religieuse. Notre société fonctionne car nous, humains, croyons en un même système de lois, nous croyons dans les valeurs de notre société, comme l'éducation ou la solidarité, nous croyons en la valeur de l'argent, alors même qu'il peut s'agir d'un simple bout de papier imprimé ou d'un nombre dans la mémoire d'un ordinateur de notre banque. En quoi la notion de croyance (ainsi définie) concerne la linguistique et la langue?

1 La langue : L'objet d'étude de la linguistique



Lectures additionnelles

Les ouvrages de Ferdinand de Saussure1916 et de Leonard Bloomfield1933 sont des monuments de la linguistique, dont la lecture reste toujours incontournable. Le livre de Wilhelm von Humboldt de 1836 est un ouvrage précurseur, mais difficile à lire aujourd'hui. La lecture des ouvrages de Claire Blanche-Benveniste est une excellente plongée dans les données véritablement attestées et ce qu'elles révèlent sur la syntaxe du français parlé et la langue en général. Le livre de Henry A. Gleason1955 est une introduction très pédagogique aux bases de la linguistique et de la syntaxe, qui a étonnamment peu vieillie. Le cours donné par Igor Mel'čuk au collège de France complétera les ouvrages de ce linguiste déjà mentionné dans l'Avant Propos. Les lecteurs intéressés par l'histoire des sciences pourront consulter les premiers travaux sur la Théorie Sens-Texte de Žolkovskij et Mel'čuk (**ŽolkovskijMel'čuk1967**), traduits en français en 1970. Les bases de la génération automatique de textes sont développées dans l'ouvrage fondateur de Laurence Danlos1987. Nous avons également mentionné les ouvrages de Steven Pinker et de Yuval Noah Harari, qui ne concernent pas directement notre sujet, mais s'inscrivent dans une approche scientifique des sciences humaines que nous suivons. Nous avons également mentionné le programme minimaliste de Noam Chomsky, présenté dans son ouvrage de 1995.

Claire Blanche-**Benveniste1990** *Le français parlé : études grammaticales*, Éditions du CNRS, Paris.

Claire Blanche-**Benveniste2000** *Approche de la langue parlée en français*, collection L'essentiel, Ophrys.

Leonard **Bloomfield1933** Language, New York (traduction française : Le langage, Payot, 1970).

Noam Chomsky1995 The minimalist program, MIT Press.

Laurence **Danlos1987** Génération automatique de textes en langues naturelles, Cambridge University Press.

Henry A. **Gleason1955** *An introduction to descriptive linguistics*, Holt, Rinehart and Wilston.

Wilhelm von **Humboldt1836** Über die Verschiedenheit des menschlichen Sprachbaus und seinen Einfluss auf die geistige Entwicklung des Menschengeschlechts (traduction en français par Pierre Caussat : Introduction à l'œuvre sur le kavi et autres essais, Seuil, 1974).

Yuval Noah **Harari2014** Sapiens : A brief history of humankind (traduction française : Sapiens : Une brève histoire de l'humanité, Albin Michel, 2015).

Mel'čuk **Igor1997** *Vers une linguistique Sens-Texte*, Leçon inaugurale, Collège de France.

Steven **Pinker1996** *How the mind works* (traduction française : *Comment fonctionne l'esprit*, Odile Jacob, 2000)

Ferdinand de Saussure1916 Cours de linguistique générale, Paris.

Aleksander Žolkovskij & Igor Mel'čuk (1967) O semanticeskom sinteze [Sur la synthèse sémantique (de textes)], *Problemy Kybernetiki* [Problèmes de Cybernétique], 19, 177-238 (traduction française : 1970, *T.A. Information*, 2, 1-85).



Citations originales

Citations de la section ??. Bloomfield (1933 : 27) :

Man utters many kinds of vocal noise and make use of the variety: under certain types of stimuli he produces certain vocal sounds and his fellows, hearing the same sounds, make the appropriate response. To put it briefly, in human speech, different sounds have different meanings. To study this co-ordination of certain sounds with certain meanings is to study language.

Citations de la section 1.3. Bloomfield1933:

Writing is not language, but merely a way of recording language by means of visible marks. [...] A language is the same no matter what system of writing may be used to record it, just as a person is the same no matter how you take is picture.

Gleason1961:

A written language is typically a reflection, independent in only limited ways, of spoken languages. As a picture of actual speech, it is inevitably imperfect and incomplete. [...] Linguistics must start with thorough investigation of spoken language before it proceeds to study written language. This is true for language with long histories of written literature, such as English, no less than those of isolated tribes which have never known of the possibility of writing.

von Humboldt1836:

Die Sprache in ihrem wirklichen Wesen aufgefasst ist etwas beständig und in jedem Augenblicke Vorübergehendes. Selbst ihre Erhaltung durch die Schrift ist immer nur eine unvollständige mumienartige Aufbewahrung die es doch erst wieder bedarf dass man dabei den lebendigen Vortrag zu versinnlichen sucht.

Citations de la section 1.5.

Bloomfield (1933:165):

By uttering a linguistic form, a speaker prompts his bearers to respond to a situation; this situation and the response to it are the linguistic meaning of the form. We assume that each linguistic form has a constant and defining meaning, different from the meaning of any other linguistic form in the same language.



Corrections des exercices

Corrigé 1 Certains linguistes s'intéressent beaucoup aux variations individuelles, notamment les chercheurs en acquisition des langues ou en socio-linguistique. Mais dans cet ouvrage, nous ne nous y intéresserons pas vraiment : nous voulons modéliser un système linguistique particulier et pas les variations entre deux systèmes. Qui plus est, nous privilégierons dans notre étude les règles qui sont communes au plus grand nombre de locuteurs. Et en même temps, nous savons qu'il existe des variations et nous voulons que le système que nous proposons soit capable de les saisir.

Corrigé 2 Non, il signifie qu'on n'est pas complètement sûr que la phrase soit grammaticale. Voir section 1.12.

Corrigé 3 Précisons pour commencer que le problème ne se situe pas du côté de la compréhension. Le problème n'est pas que cette phrase ne sera pas comprise, le problème est qu'elle ne pourra pas être produite par un locuteur dans une situation normale de communication. Pour la produire, nous avons fait un travail de linguiste et pas de locuteur ordinaire, en appliquant des insertions successives de relatives. À partir de là, nous considérons que les locuteurs n'ont pas la compétence de produire de tels énoncés et que cela n'a rien à voir avec la performance. Mais on peut aussi estimer (comme Noam Chomsky) que cette limitation n'est pas à proprement parler une propriété de la langue et qu'elle relève d'une modélisation des capacités cognitives en général. Si le modèle linguistique est combiné avec un tel modèle, on peut ne pas tenir compte des limitations mémorielles dans la modélisation des langues. Quoi qu'il en soit, il faut, à notre avis, les prendre en compte quelque part dans un modèle de la compétence.

Corrigé 4 Ces deux énoncés présentent une parenthèse (je sais pas pourquoi dans le premier et je sais pas à quel niveau c'est exactement dans le

second). Il s'agit d'un énoncé syntaxiquement indépendant à l'intérieur d'un autre énoncé. Cela illustre la capacité du locuteur à commenter ce qu'il dit (comme si quelqu'un d'autre venait interrompre son propos) sans pour autant perdre le fil de ce qu'il dit.

Toute langue repose sur un système de conventions que les locuteurs doivent accepter. De la même facon que nous acceptons qu'un billet de 20€ a pour tous nos condisciples deux fois plus de valeur qu'un billet de 10€, nous acceptons que le texte *un chien* désignera bien un chien pour nos condisciples. L'apprentissage d'une langue repose donc sur la croyance que les conventions linguistiques seront acceptées par nos condisciples. Pour qu'une langue existe, il faut qu'un groupe d'humains accepte de croire au système de conventions qu'elle suppose. On peut citer **Saussure1916**: « [La langue] est la partie sociale du langage, extérieure à l'individu, qui à lui seul ne peut ni la créer ni la modifier; elle n'existe qu'en vertu d'une sorte de contrat passé entre les membre de la communauté. » On considère généralement qu'il existe deux types d'entités : les entités objectives, qui appartiennent au monde réel, et les entités subjectives, que je crée dans mon cerveau. Il existe en fait un troisième type d'entités, qu'Harari appelle les entités inter-subjectives, et qui n'existent qu'à travers un accord entre les individus d'une communauté : Zeus, l'euro (la monnaie), la France, Google (la société) ou encore la langue française. Toutes ces choses n'existent que parce que les humains pensent qu'elles existent. Si les humains disparaissent, ces choses disparaîtront avec eux.

2 Produire un énoncé : La syntaxe mise en évidence par un exemple

2.1 Analyse et synthèse

Ce chapitre est entièrement consacré à l'étude de la production d'un énoncé ou plus exactement à la production d'une famille d'énoncés concurrents exprimant plus ou moins le même sens. Il est habituel de commencer l'étude d'une langue en analysant les textes (éventuellement oraux) produits dans cette langue. L'étude procède alors dans le sens de l'ANALYSE, c'est-à-dire du texte vers le sens. Nous pensons, à la suite de Lucien Tesnière et d'Igor Mel'čuk, qu'il est préférable d'étudier la langue dans le sens de la SYNTHÈSE, c'est-à-dire du sens vers le texte.

Pour comprendre le fonctionnement de la langue, il est donc nécessaire d'étudier la façon dont un énoncé est produit par un locuteur et les opérations qui conduisent à la production de cet énoncé. Nous mettrons en évidence un ensemble de contraintes auxquelles doit obéir le locuteur lors de l'énonciation et qui constitue la grammaire de la langue. Parmi ces contraintes, nous verrons que la langue nous impose une structuration hiérarchique qui constitue le cœur de la syntaxe.

2.2 Les observables : textes et sens

Un modèle se construit à partir d'OBSERVABLES, puisqu'il modélise le fonctionnement de quelque chose de réel, que l'on peut observer. Néanmoins, un modèle est amené à faire des hypothèses sur la façon dont ces observables sont produits et à construire ainsi un certain nombre d'objets qui ne sont pas observables. Pour ce qui est de la langue, deux choses sont réellement observables : les textes et les sens. Pour les textes, c'est assez évident, même si la question de savoir quelle est la représentation d'un texte dans notre cerveau reste un problème. Pour les sens, c'est plus complexe : on ne peut pas observer le sens en tant que tel, mais on peut s'assurer qu'un texte est compréhensible et qu'il est compris. L'un des outils d'observation est la paraphrase : on peut demander à un locuteur de

reformuler un texte ou lui demander si deux textes ont le même sens. On peut ainsi considérer, à la suite d'Igor Mel'čuk, que « avoir le même sens » est un observable et **définir le sens** comme un invariant de paraphrases, c'est-à-dire comme ce qui est commun à tous les textes qui ont le même sens.

Nous faisons l'hypothèse qu'il y a entre le sens et le texte un niveau d'organisation syntaxique. Cette organisation n'est pas directement observable. Dans la suite, nous serons amenés à construire un grand nombre d'objets linguistiques et notamment des représentations syntaxiques. Il s'agit de **constructions théoriques** et de rien d'autre. On ne prouve leur « existence » qu'à l'intérieur d'une théorie. Elles existent dans le modèle, mais cela ne prouve pas qu'elles existent dans l'objet réel que nous modélisons. On ne peut pas réfuter directement leur existence. On ne peut que réfuter l'adéquation du modèle avec l'objet modélisé. Néanmoins, si le modèle est adéquat et suffisamment économique, alors les objets construits par le modèle acquièrent une part de réalité, deviennent plus tangibles.

2.3 Partir d'un sens

Considérons un locuteur du français qui veut exprimer un certain sens que nous allons essayer de décrire sans le verbaliser directement par une phrase. Ce sens concerne une personne x appelée Ali et un événement e concernant e. Cet événement est une maladie et la durée de cette maladie est de deux semaines. Nous pouvons représenter ce sens par la « formule » suivante :

x: 'Ali'
e: 'malade'(x)
'durer'(e,'2 semaines')

Figure 2.1 – Description formelle d'un sens

Dans cette représentation, il y a quatre éléments de sens considérés : 'Ali', 'malade', 'durer' et '2 semaines'. Ce dernier élément est en fait la combinaison de deux sens : l'unité de temps 'semaine' et le prédicat 'deux' qui la quantifie. Ces éléments de sens sont des sens de mots du français. Ils peuvent être exprimés de façons variées : par exemple, le sens 'durer' peut aussi bien être exprimé par le nom durée que le verbe durer ou encore la préposition pendant comme nous le verrons plus loin. Remarquons que nous distinguons les unités lexicales, notées en majuscule (durée), de leur sens, noté entre guillemets simples ('durer'). Un sens exprimable par une unité lexicale est appelé un sens lexical. On utilise

parfois le terme *sémantème* pour désigner un sens lexical, mais nous réserverons ce terme pour des signes linguistiques élémentaires (voir chapitres 2.3 et 2.4).

Certains sens lexicaux fonctionnent comme des prédicats et possèdent des arguments : le sens 'malade' possède toujours un argument (*quelqu'un* est malade), le sens 'durer' possède deux arguments (*quelque chose dure quelque temps*). Le sens 'Ali', quant à lui, renvoie à une entité du monde (ou plus exactement à la représentation mentale qu'en a le locuteur) et n'a pas d'argument. Dans la formule ci-dessus, les variables x et e nous ont servi à indiquer que certains sens étaient arguments d'autres : x désigne le sens 'Ali' et 'malade'(x) signifie que l'élément désigné par x est l'argument de 'malade'.

Les relations prédicat-argument entre les sens lexicaux constituent la structure prédicative. La structure prédicative exprime le contenu informationnel.

Nous nous contenterons de cette définition sommaire de la structure prédicative (voir compléments dans l'encadré ?? sur Les composantes du sens). Une définition plus précise sera donnée au Chapitre ?? sur La structure syntaxique profonde.

Nous avons présenté la structure prédicative par une formule, mais on peut aussi la représenter par un graphe (voir l'encadré ?? qui suit) que nous appelons le graphe sémantique. Les nœuds du graphe sémantique sont les sens lexicaux 'Ali', 'malade', 'durer', 'semaine' et 'deux' et les arêtes représentent les relations prédicat-argument. Les arêtes sont matérialisées par des flèches. Les arêtes sont étiquetées par des chiffres permettant de distinguer les différents arguments d'un prédicat : 1 pour le premier argument, 2 pour le deuxième, etc. L'ordre dans lequel les arguments sont numérotés est l'ordre de saillance, dont nous reparlerons au Chapitre ?? sur Les relations syntaxiques.

Figure 2.2

2.4 Graphe et arbre 🕏



Graphe et arbre sont des structures mathématiques très utilisées en sciences. Elles sont utilisées en linguistique pour la représentation du sens et de la structure syntaxique.

Un graphe est une structure liant ensemble des éléments. Les éléments sont appelés les NŒUDS du graphe et les liens les ARÊTES. Un graphe est dit CONNEXE lorsque pour chaque couple de nœuds du graphe, il existe un ensemble d'arêtes formant un chemin connectant un nœud à l'autre.

Graphe connexe Graphe non connexe

Un graphe est dit ACYCLIQUE s'il n'existe aucun chemin partant d'un nœud et revenant à ce nœud sans emprunter deux fois la même arête.

Graphe acyclique Graphe avec cycle

Un arbre est un graphe connexe et acyclique dont un nœud, qu'on appelle la racine, est pointé. Cela revient à orienter les arêtes à partir de ce point. On peut donc aussi définir un graphe comme un cas particulier de graphe orienté.

Un GRAPHE ORIENTÉ est un graphe dont les arêtes sont orientée, c'est-à-dire distinguent un nœud source et un nœud cible de l'arête. On représente généralement l'orientation par une flèche allant de la source à la cible.

Un arbre est donc aussi un graphe orienté connexe pour lequel chaque nœud est la cible d'une seule arête à l'exception d'un nœud, la racine de l'arbre. Tout nœud autre que la racine de l'arbre possède ainsi un gouverneur qui est l'unique nœud qui le prend pour cible. On représente traditionnellement les arbres « à l'envers » avec la racine en haut. Les nœuds qui ne sont le gouverneur d'aucun nœud, c'est-à-dire qui n'ont pas de dépendants, sont appelés les feuilles de l'arbre. Un chemin orienté allant de la racine ou d'un nœud intérieur à une feuille est appelé une branche. Du fait que, par convention, chaque arête est toujours orientée vers le bas (la cible est en dessous de la source), il n'est pas nécessaire d'utiliser une deuxième convention et d'indiquer l'orientation par une flèche.

Dépendance Branche

La structure syntaxique d'une phrase est généralement représentée par un arbre dont les nœuds sont les unités lexicales.

Notons encore qu'il existe une notion d'acyclicité plus restrictive pour les graphes orientés : un graphe orienté est dit ACYCLIQUE s'il n'existe aucun chemin orienté permettant de partir d'un nœud et de source en cible de revenir au même nœud.

Graphes avec cycle orienté Graphe sans cycle orienté

La **structure prédicative** d'un énoncé peut être formalisée par un **graphe orienté**



connexe et acyclique (encore appelé DAG, de l'anglais directed acyclic graph) dont les nœuds sont les sens lexicaux et grammaticaux et les arêtes sont les dépendances sémantiques ou relations prédicat-argument entre ces sens.

2.5 Les composantes du sens



Nous avons déjà expliqué dans le chapitre précédent la distinction que nous faisons entre le sens linguistique et les intentions communicatives du locuteur. Reste à savoir, bien que cette question dépasse le cadre de cet ouvrage consacré à la syntaxe, comment modéliser le sens linguistique. Nous pensons que la structure prédicative (voir la section 2.3 Partir d'un sens) exprime une partie et une partie seulement du sens linguistique : il s'agit de l'information pure contenue dans le message – quels sont les sens de base que nous souhaitons utiliser et comment ils se combinent.

Au contenu informationnel s'ajoute au moins quatre autres types de contenus :

- La STRUCTURE COMMUNICATIVE, dont nous discuterons plus loin dans ce chapitre, indique ce qui est réellement informatif pour le destinataire, c'està-dire ce qu'on suppose qu'il sait déjà, ce qu'on souhaite souligner ou au contraire mettre en arrière-plan, etc. On appelle aussi cette structure l'emballage DE L'INFORMATION, de l'anglais information packaging. Le terme le plus employé actuellement est structure informationnelle, de l'anglais information structure, mais nous éviterons absolument ce terme qui est une source de confusion évidente avec le contenu informationnel.
- La structure rhétorique indique le style (familier, poétique, humoristique, etc.) avec laquelle l'information sera communiquée (voir section suivante).
- La STRUCTURE ÉMOTIONNELLE indique quelles sont les émotions liées à cette information. Elle a surtout un impact sur la prosodie, mais peut influencer certains choix lexicaux (des termes injurieux par exemple).
- Par ailleurs, le contenu informationnel ne nous informe que si nous pouvons l'ancrer dans la réalité (ou plus exactement la représentation du monde que les interlocuteurs construisent dans leur cerveau à partir de la réalité), c'est-à-dire si on peut décider, par exemple, si 'Ali' renvoie à un objet du monde que nous connaissons déjà ou pas et si oui lequel. Les liens entre le contenu informationnel et le monde (ou plus précisément le CONTEXTE D'ÉNONCIATION, c'est-à-dire la partie du monde concernée par l'énonciation) constitue la structure référentielle, La référence joue surtout un rôle dans la planification et le choix de l'information permettant de s'assu-

rer que l'interlocuteur identifiera le bon référent, mais une fois le message élaboré, la structure référentielle a peu d'incidence sur la réalisation du message. Ce qui compte surtout, c'est si le locuteur présente l'information comme nouvelle ou non et ceci appartient à la structure communicative.

Ce découpage du sens (à l'exception de la structure émotionnelle) et la représentation du contenu informationnel par un graphe sémantique ont été proposés par Žolkovskij et Mel'čuk en 1965 dans leur introduction à la Théorie Sens-Texte.

Notons encore qu'il existe des relations d'équivalence entre les structures prédicatives : une configuration de sens lexicaux peut être remplacée par un seul sens ou une autre configuration sans modifier le sens global. Un sens lexical peut être ainsi décomposé à la manière de ce que l'on fait quand on donne une définition d'un des sens d'un mot dans un dictionnaire. On peut considérer, à la suite d'Anna Wierzbicka (*Lingua Mentalis*, Academic Press, 1980), qu'il existe un ensemble d'unités minimales de sens à partir desquelles peuvent être définis tous les autres sens lexicaux. (De tels ensembles contenant une cinquantaine de sens minimaux ont été proposés.) Un contenu informationnel correspond ainsi à un ensemble de structures prédicatives équivalentes, dont les sens lexicaux sont plus ou moins décomposées.

2.6 Choisir des unités lexicales

Comment peut-on exprimer le sens que nous avons considéré en français? Nous allons nous intéresser à deux formulations possibles de ce sens, suffisamment différentes pour illustrer notre propos.

- (1) La maladie d'Ali a duré deux semaines.
- (2) Ali a été malade pendant deux semaines.

L'analyse de la production de ces deux énoncés va nous permettre de mettre en évidence de nombreuses règles de grammaire, appartenant à la langue en général ou spécifiques au français.

La première chose qui va déterminer la nature du texte que nous produisons est la façon dont nous réalisons chacun des éléments de sens du message que nous souhaitons communiquer. Pour simplifier, nous considérons que chaque sens va être réalisé par une unité lexicale : par exemple, 'malade' peut être réalisé par l'adjectif MALADE ou le nom MALADIE; 'durer' peut être réalisé par le nom DURÉE, le verbe durer ou les prépositions durant ou pendant.



Q

Comment se font les CHOIX LEXICAUX? Les choix lexicaux dépendent bien sûr du sens que l'on veut exprimer : par exemple, pour parler de l'ingestion d'un aliment, on devra choisir entre BOIRE et MANGER selon la nature de cet aliment. mais aussi entre manger, déguster ou dévorer selon la façon dont on l'a ingéré ou encore entre manger et bouffer selon la familiarité avec laquelle nous avons l'habitude de nous adresser à notre interlocuteur. Cependant, la subtilité du sens que nous voulons exprimer n'est pas le seul facteur qui contraint les choix lexicaux. Ils existent d'autres contraintes qui relèvent de la grammaire et dont nous allons parler maintenant.

2.7 Les quatre moyens d'expression du langage



Lorsque nous parlons, nous avons quatre moyens à notre disposition pour exprimer du sens.

- Le premier moyen, ce sont les **mots** ou plus exactement les **unités lexicales** et grammaticales (voir la partie 2 consacrée aux Unités de la langue).
- Le deuxième moyen, c'est la façon de combiner les mots et en particulier l'**ordre** dans lequel nous les mettons : « *Ali regarde Zoé* » ne veut pas dire la même chose que « Zoé regarde Ali ». Plus subtilement, « À Paris, Ali travaille le lundi » n'est pas exactement synonyme de « Le lundi, Ali travaille à Paris » (le premier énoncé n'implique pas qu'Ali travaille tous les lundis et qu'il est à Paris tous les lundis où il travaille).
- Le troisième moyen d'expression du langage est la prosodie. La séquence tu viens peut être prononcée de bien des façons et avoir autant de sens différents. Prononcée avec une voix forte et autoritaire et un accent montant sur la première syllabe, ce sera un ordre : « TU VIENS! ». Prononcée d'une voix suave avec une courbe mélodique montante, ce sera une question ou une invitation : « *Tu viens*? ». Prononcée d'une voix neutre avec une courbe mélodique descendante ce sera une simple constatation : « Tu viens. ».
- Le quatrième moyen à notre disposition dépasse le simple usage de la voix. Une énonciation en face à face s'accompagne toujours de mimiques faciales et de gestes divers. Ceux-ci sont beaucoup plus codifiés et beaucoup plus riches qu'on ne le pense généralement. Ils vont accompagner la parole et parfois se combiner avec elle. Ainsi « Ton pull » suivi d'un geste avec le pouce levé est un énoncé équivalent à « Ton pull est vraiment super ». Et une moue désapprobatrice en prononçant « *fean*? » en dira beaucoup plus qu'un long discours sur la confiance qu'on met en Jean.

2.8 Contraintes syntaxiques sur les choix lexicaux

À peu près n'importe quel énoncé en français possède un verbe principal et ce verbe est conjugué. Cela signifie que l'un des sens de notre message devra être lexicalisé par un verbe, par exemple 'durer' par durer. L'autre possibilité est de lexicaliser 'malade' par un verbe, mais comme il n'existe pas de verbe *malader en français, nous lexicalisons 'malade' par un adjectif et nous en faisons une tournure verbale être malade grâce au verbe être. Le verbe être n'a donc aucune contribution sémantique ici; il a juste un rôle grammatical, qui est d'assurer que l'élément principal de la phrase est bien un verbe et donc de faire d'un adjectif l'équivalent d'un verbe. Ce rôle très particulier du verbe être lui vaut le nom de copule.

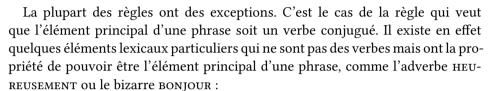
Une fois choisi l'élément principal de la phrase, les éléments lexicaux choisis ensuite se voient imposer un certain nombre de choses, à commencer par leur PARTIE DU DISCOURS (les principales parties du discours du français sont nom, verbe, adjectif et adverbe). L'élément principal de la phrase est un verbe conjugué comme on vient de le dire (ceci sera justifié dans le chapitre consacré à la tête dans la Partie 3). Les éléments qui vont dépendre de ce verbe devront ensuite être soit des noms, soit des adverbes selon la relation qu'ils entretiennent avec ce verbe.

Commençons par le cas de la phrase (7) (La maladie d'Ali a duré deux semaines) dont le verbe principal est durer : le sens 'malade', qui est le premier argument du sens 'durer' devra être réalisé comme le sujet de durer et devra être un nom. La lexicalisation de 'malade' par un nom donne ainsi MALADIE. Le sens 'semaine' sera également réalisé par un nom, qui sera un complément direct du verbe. Nous discuterons dans le chapitre sur les fonctions syntaxiques de la Partie 6 de ces compléments de mesure qui ressemblent à des compléments d'objet directs mais n'en possède pas toutes les bonnes propriétés.

Le cas de la phrase (8) (Ali a été malade pendant deux semaines) est plus complexe. Son élément principal est la tournure verbale être malade. L'unique argument de 'malade' est 'Ali', qui sera donc réalisé comme sujet de la tournure verbale et sera un nom, Ali en l'occurrence. Le sens 'durer' est lui aussi directement lié à 'malade' et devra donc être réalisé comme un dépendant direct de la tournure verbale. Mais contrairement aux cas précédents, 'durer' n'est pas un argument de l'élément principal : c'est même l'inverse, c'est lui qui prend l'élément principal comme argument sémantique. Dans un tel cas, l'élément de sens doit être réalisé par un groupe adverbial. C'est ainsi que le sens 'durer' peut être réalisé par la préposition PENDANT. Le deuxième argument de 'durer' est réalisé

par un nom, qui est le complément de la préposition. La préposition et son complément, pendant deux semaines, forme un groupe de distribution équivalente à un adverbe (comme *longtemps* par exemple).

2.9 Règles et exceptions



Heureusement qu'il y en a!

Sinon, bonjour le chômage des linguistes.

Par ailleurs, il existe des contextes où la règle ne s'applique pas, notamment en réponse à une question (Tu viens à la fac demain? Oui à 10h pour le cours de syntaxe) ou encore pour les titres (Nouvel incident diplomatique entre la France et l'Allemagne). Cela ne signifie pas que la règle est fausse, mais qu'il faut bien préciser quand elle s'applique. Quant aux exceptions, il faut les lister et inclure ces éléments dans une classe particulière d'élément que nous appelons les prédicatifs et les locutifs (voir le Chapitre ?? sur Les catégories microsyntaxiques).

2.10 Structure hiérarchique

Lors du passage du sens au texte, tout se passe comme si on suspendait le graphe sémantique par l'un de ses nœuds dont on décide de faire le verbe principal de l'énoncé et qu'on parcourait le graphe à partir de ce nœud pour les autres lexicalisations. Nous pouvons illustrer cela par les schémas suivants : à gauche nous représentons le graphe sémantique suspendu par un de ses nœuds (visé ici par une grosse flèche blanche) et à droite nous avons la structure hiérarchique correspondante où les sens lexicaux ont été lexicalisés.

Figure 2.3 – Hiérarchisation à partir de 'durer'

Figure 2.4 - Hiérarchisation à partir de 'malade'

La structure que nous obtenons s'appelle un arbre de dépendance syntaxique PROFOND (voir le Chapitre ?? entièrement consacré à La syntaxe profonde). Chaque nœud de l'arbre est occupé par une unité lexicale correspondant à un sens. D'autres sens (non considéré ici) donne des éléments grammaticaux qui vont se combiner aux unités lexicales : temps pour les verbes (présent, passé, etc.), nombre (singulier, pluriel) et définitude (défini, indéfini) pour les noms. Le nœud au sommet est appelé la RACINE de l'arbre (voir l'encadré ?? sur *Graphe et arbre*). Tous les nœuds à l'exception de la racine dépendent d'un autre nœud appelé leur gouverneur (syntaxique profond). À l'inverse, les nœuds qui dépendent d'un autre nœud en sont appelés les dépendants (syntaxiques profonds). Le lien entre deux nœuds est appelé une dépendance (syntaxique profonde).

Chaque dépendance est étiquetée en fonction de son parcours : les relations sémantiques qui ont été parcourues dans le sens de la flèche (du prédicat vers l'argument) donne une DÉPENDANCE SYNTAXIQUE ACTANCIELLE (que nous numérotons comme la dépendance sémantique correspondante), tandis que les relations sémantiques qui ont été parcourues à contre-courant donne une DÉPENDANCE SYNTAXIQUE MODIFICATIVE (étiquetée *MOD*).

Dépendance actancielle Dépendance modificative

Figure 2.5

Comme nous l'avons expliqué dans la section précédente, chaque unité lexicale se voit imposer sa partie du discours par son gouverneur, la racine étant un verbe conjugué. Ainsi dans la phrase (7), ALI, qui dépend du nom MALADIE, devrait être réalisé par un adjectif, ce que nous indiquons par la notation $Adj_0(ALI)$. C'est la préposition DE qui assurera cette Translation ($la\ maladie\ d'Ali$) et permettra à ALI d'être complément du nom (voir la section du Chapitre ?? sur $La\ translation$). De la même façon, la notation $V_0(MALADE)$ indique que MALADE occupe une position où un verbe est attendu et c'est la copule ÊTRE qui assurera la translation d'adjectif en verbe.

2.11 Du sens au texte : de 3D à 1D 🍳

Nous avons mentionné dans la section 1.3 Sons et textes le caractère unidimensionnel de la chaîne parlée. Le sens lui est localisé dans notre cerveau : un certain nombre de zones de notre cerveau s'activent simultanément (ou les unes après les autres) et se mettent en réseau : le sens est donc fondamentalement un objet au moins tridimensionnel (quadridimensionnel si l'on prend en compte la dimension temporelle et le caractère dynamique de la construction du sens). Le passage du sens au texte s'accompagne donc d'une réduction de dimensionnalité, un passage de la dimension 3 à la dimension 1, une suite ordonnée d'unités élémentaires.

Il semble que la langue effectue ce changement de dimension en deux étapes :

- le passage de la dimension 3 à la dimension 2 est une hiérarchisation du sens, le passage d'un graphe à un arbre;
- le passage de la dimension 2 à la dimension 1 est la linéarisation de ce graphe.

La phase de hiérarchisation s'accompagne d'une phase de « lexicalisation », c'est-à-dire de choix de signes linguistiques élémentaires, des unités lexicales et des unités grammaticales, qui se contraignent les unes les autres. La phase de linéarisation s'accompagne d'une « morphologisation » des signes, c'est-à-dire de combinaison des signifiants selon des règles morpho-phonologiques propres à chaque langue. Une telle architecture est à la base de la Théorie Sens-Texte, que nous évoquerons dans l'encadré ?? éponyme.

sens
hiérarchisation
« lexicalisation »
arbre syntaxique
linéarisation
« morphologisation »
chaîne parlée

2.12 Ce que la langue nous force à dire

En produisant les phrases (7) et (8) (section 2.6), nous avons exprimé plus que le sens de départ. Celui-ci ne contenait pas d'information temporelle sur le moment de la maladie d'Ali, mais nous avions besoin d'une telle information pour conjuguer le verbe principal. C'est la grammaire du français qui nous contraint, en nous obligeant à ajouter une flexion au verbe principal de la phrase, à situer l'événement dont nous voulons parler par rapport au moment où nous parlons. Nous devons décider si la maladie a eu lieu avant maintenant (Ali a été malade pendant deux semaines) ou si l'événement aura lieu après maintenant (Ali sera malade pendant deux semaines) ou encore s'il est en cours (Ali est malade depuis deux semaines).

De même, en français, on ne peut pas lexicaliser un sens par un nom sans préciser le nombre (*Ali a mangé une pomme* vs *Ali a mangé des pommes*) et sans préciser si la chose est déjà connue ou non (*Ali a mangé la pomme* vs *Ali a mangé une pomme*) (voir Chapitre ??). C'est l'une des raisons pour lesquelles nos deux phrases de départ (7) et (8) ne sont pas parfaitement synonymes, puisque, dans (7), 'malade' est lexicalisé par un nom et est donc accompagné d'un article défini,

lequel présuppose que la maladie d'Ali est connue des interlocuteurs au moment où la phrase est prononcée, ce qui n'est pas le cas en (8).

En conclusion, il n'est pas possible en français de communiquer uniquement notre sens de départ! Nous devons y ajouter des informations et en particulier situer le fait dont nous parlons (la maladie d'Ali) dans le temps.

2.13 Les sens grammaticaux 🕏

Dans un article de 1959 sur la traduction, le grand linguiste d'origine russe Roman Jakobson remarque que « Les langues diffèrent essentiellement par ce qu'elles *doivent* communiquer et pas par ce qu'elles *peuvent* communiquer. » Il explique par exemple que la traduction de la phrase anglaise *I hired a worker* en russe nécessiterait deux informations supplémentaires : le verbe en russe devra indiquer si l'embauche a été complétée ou pas, tandis que le nom en russe devra indiquer le genre et donc s'il s'agit d'un travailleur ou d'une travailleuse. À l'inverse, la phrase russe n'aura pas à choisir entre article défini ou indéfini (*une* ou *la travailleuse*), ni entre les temps verbaux *hired* vs. *have hired*. Sur cette nécessité de faire des choix dès qu'il s'agit de catégories grammaticales obligatoires, Jakobson renvoie à Franz **Boas1938** : « Nous devons choisir parmi ces aspects, et l'un ou l'autre doit être choisi. »

Il existe quantités de sens plus ou moins curieux qui sont ainsi imposés par la grammaire d'une langue. Par exemple, il existe une langue amérindienne, le nootka, parlée sur l'île de Vancouver au Canada, où le verbe s'« accorde » avec son sujet en fonction de particularités physiques du référent sujet; on doit nécessairement choisir entre l'une des sept possibilités suivantes : normal, trop gros, trop petit, borgne, bossu, boiteux, gaucher (Sapir1915; Mel'čuk1993). On ne peut donc pas dire ce que fait une personne donnée sans dire si cette personne est normale ou si elle est affublée d'une des six particularités physiques retenues par la grammaire de cette langue.

Si la nature des sens qui sont grammaticalement exprimés au travers des langues du monde est assez vaste, elle est quand même assez homogène. D'une part, il s'agit généralement de sens abstraits ayant une **importance cognitive fondamentale** et qui ont de surcroît une importance culturelle énorme, puisqu'ils apparaissent dans pratiquement chaque phrase et donc modèlent la pensée des locuteurs à chaque instant. D'autre part, ces sens s'associent naturellement à certaines parties du discours : les langues ont ainsi généralement une classe d'éléments lexicaux qui varient en temps et/ou aspect et désignent des procès (et correspondent grosso modo à nos verbes) et une classe d'éléments lexicaux qui

varient en nombre et désignent des entités (et correspondent à nos noms).

On trouvera dans le volume 2 du *Cours de morphologie générale* d'Igor Mel'čuk publié en 1993 une typologie de tous les sens qui doivent être exprimés obligatoirement dans une langue au moins. Nous donnons ici l'exemple du **respect** en japonais.

Un Japonais ne peut pas dire « *Pierre est malade* » sans préciser deux choses qui, au premier abord, peuvent sembler étonnantes à un locuteur du français :

- est-ce que Pierre est quelqu'un de respectable ou non?
- est-ce que la personne à qui je parle est quelqu'un de respectable ou non?
 Ces deux informations doivent obligatoirement figurer dans le choix lexical de l'adjectif et la conjugaison de la copule :
 - 1. Mannana Pieru-san-wa go-byoki desu.
 - 2. DDDDDDDDD Pieru-wa byoki **desu**.
 - 3. ⊠□□□□□□□⊠ Pieru-san-wa go-byoki da.
 - 4. □□□□□□□□ Pieru-wa byoki da.

La respectabilité envers Pierre est exprimée par le suffixe san et par le choix d'une forme polie GO-BYOKI de l'adjectif 'malade'. La respectabilité envers l'interlocuteur est exprimée par la forme polie de la copule, *desu*, opposée à la forme neutre *da*. Ces deux marques de respectabilité sont indépendantes, ce qui nous donne quatre formes possibles.

La respectabilité envers l'interlocuteur peut être comparée au **vouvoiement** en français. Mais alors que le choix entre *tu* et *vous* se limite au cas, évitable, où l'on interpelle directement son interlocuteur, le choix entre la forme verbale respectueuse et la forme familière se pose pour chaque phrase en japonais.

2.14 Choisir le verbe principal

Nous savons que la structure syntaxique est une structure hiérarchique et que la construction de cette structure commence par le choix d'un élément sémantique pour être la racine de l'arbre syntaxique et donc le verbe principal de l'énoncé produit. Le choix de l'élément principal de la phrase peut être guidé par les choix lexicaux (quel sens lexical peut donner un verbe) ou grammaticaux (sur quel sens lexical veut-on ou peut-on ajouter tel ou tel sens grammatical). Mais ce choix est avant tout guidé par ce dont on est en train de parler et ce qu'on veut dire. Ainsi le même message, selon que nous sommes en train de parler d'Ali ou de sa maladie sera exprimé différemment : si l'on est en train de parler d'Ali et

que l'information nouvelle que l'on veut communiquer est sa maladie, on choisira plutôt (8) (Ali a été malade pendant deux semaines), alors que si l'on est en train de parler de sa maladie et que l'information nouvelle que l'on veut communiquer est seulement sa durée, on peut préférer (7) (La maladie d'Ali a duré deux semaines). Le verbe principal de la phrase est ainsi généralement l'élément central du RHÈME, c'est-à-dire l'élément de sens que l'on souhaite prioritairement communiquer, tandis que le sujet du verbe principal est généralement le THÈME, c'est-à-dire ce dont on parle, ce sur quoi porte le rhème. L'indication des rhème et thème ne fait pas partie du contenu informationnel, représenté par la structure prédicative, mais dépend de la façon dont on communique le contenu informationnel, ce que nous avons appelé la structure communicative (voir l'encadré ?? sur Les composantes du sens). Formellement, la délimitation des thème et rhème se surajoute à la structure prédicative en indiquant quelle zone constitue le rhème et quelle autre le thème.

2.15 Les contraintes de la grammaire

Reprenons rapidement la liste des contraintes qui nous ont été imposées lors de la production de nos deux phrases. La principale contrainte est la structure hiérarchique de l'énoncé. De cette contrainte majeure découlent des contraintes sur les parties du discours des unités lexicales choisies : la racine de l'arbre syntaxique devra être un verbe, les actants d'un verbe des noms, les modifieurs d'un verbe des adverbes et les dépendants d'un nom des adjectifs (pour les notions d'actant et de modifieur, voir le Chapitre ?? sur la *Syntaxe profonde*). Ainsi les sens 'malade' et 'durer' peuvent recevoir différente lexicalisations (MALADE vs MALADIE, DURER vs PENDANT), mais ces différents choix ne sont pas totalement indépendants : il serait par exemple difficile de faire une phrase naturelle avec les unités lexicales malade et durer exprimant notre sens de départ. (On peut en faisant deux phrases : « *Pierre a été malade. Ça a duré deux semaines.* ».)

Quand les unités lexicales n'appartiennent pas à la partie du discours attendues, d'autres unités lexicales sont introduites, comme la copule ÊTRE ou la préposition DE, pour « masquer » le mauvais choix catégoriel. En fonction de leur partie du discours, les unités lexicales se voient assignées des unités grammaticales, telles que le temps pour les verbes et le nombre et la définitude pour les noms. Il existe encore d'autres contraintes. Par exemple, en français, un verbe conjugué, à l'exception de l'impératif, doit toujours avoir un sujet et ce sujet devra se placer devant lui (il existe quelques cas où le sujet peut être « inversé » mais pas ici). Enfin, le verbe devra s'accorder avec son sujet. De même, le nom



devra avoir un déterminant (article ou autre) exprimant la définitude, accordé en nombre avec le nom placé avant lui. Ces différentes contraintes – la nature de l'élément principal d'une phrase, la partie du discours des dépendants, les unités grammaticales obligatoires comme la conjugaison pour les verbes ou l'article pour les noms, l'accord, le placement des mots les uns par rapport aux autres -, sont des règles grammaticales du français. La grammaire d'une langue, ce sont toutes les contraintes que nous impose cette langue quand nous parlons.

L'expression même des contraintes grammaticales repose sur les « connexions » entre différentes parties de l'énoncé. Ces connexions constituent, comme nous le verrons, le squelette des relations syntaxiques entre les mots, les groupes de mots, mais aussi des unités plus petites que les mots. En mettant en évidence les contraintes que ces unités, que nous appellerons les UNITÉS SYNTAXIQUES, s'imposent les unes aux autres, nous avons mis en évidence l'existence d'une STRUCTURE SYNTAXIQUE. L'étude de cette structure, qui modélise la façon dont les signes linguistiques se combinent les uns avec les autres, constitue l'objet central de la SYNTAXE et donc de cet ouvrage.

2.16 D'où viennent les règles de la grammaire?



D'où viennent ces règles? Ces règles n'ont pas été inventées par des grammairiens; elles existent depuis beaucoup plus longtemps que les grammairiens et elles sont pour la plupart indépendantes d'eux. Aussi le terme « régularité » correspondrait-il peut-être mieux que « règle » au caractère « naturel » de ces contraintes imposées par la grammaire de la langue. Le linguiste cherche juste à décrire la grammaire qu'utilisent naturellement les locuteurs natifs d'une langue.

Il nous faut quand même dire quelques mots de la GRAMMAIRE NORMATIVE, élaborée par les « grammairiens » et dont l'objectif est de prescrire le « bon » usage du français. L'institutionnalisation de la normativité n'est pas universelle, mais plutôt une particularité française. Dans des traditions différentes, en Allemagne ou aux États-Unis par exemple, on considère davantage qu'en France qu'un locuteur natif sait parler sa langue : on n'enseigne que très peu de grammaire à l'école et il n'y a pas d'Académie nationale chargée de protéger la langue.

Quel est le bon français?

La distinction entre grammaire descriptive et grammaire normative est importante quand on tente de répondre à la question suivante : la phrase suivante est-elle grammaticale?

Après qu'il se soit assis, elle s'est mise à parler.

La réponse est non selon les grammaires normatives qui estiment que APRÈS QUE doit être suivi de l'indicatif, et oui, si on se base sur une grammaire descriptive, étant donné qu'il s'agit d'un énoncé contenant des structures attestées régulièrement. Une grammaire descriptive nous informera qu'en français contemporain, la conjonction APRÈS QUE est suivie soit de l'indicatif, soit du subjonctif. La description tentera peut-être d'expliquer quand on utilise l'un ou l'autre. De plus, elle notera probablement que l'utilisation de l'indicatif a un caractère plus écrit à cause de la prescription de la grammaire normative.

Pour la grammaire normative seule l'utilisation de l'indicatif est correcte. Pour ce jugement, les grammaires normatives se basent sur :

- un certain conservatisme : « ce qui est vieux est meilleur »;
- et sur un certain élitisme : « la langue des riches, puissants et instruits est meilleure ».

La linguistique ne s'intéresse pas à ces jugements esthétiques. Les règles linguistiques sont toujours descriptives et jamais prescriptives.

Quand les grammairiens ont gagné

La grammaire normative peut avoir une influence sur le comportement linguistique des locuteurs d'une langue. En effet, une prescription peut être acceptée par la langue et donc entrer dans les régularités de la langue. Dans ce cas-là, la linguistique s'intéresse bien entendu à cette régularité.

Un exemple de ce phénomène concerne l'accord au masculin. Aujourd'hui, les francophones accordent naturellement au masculin des éléments coordonnés composés de noms de genre différents. On dit donc :

(4) Arrivèrent alors un homme et cinquante femmes très élégants.

et non:

(5) *Arrivèrent alors un homme et cinquante femmes très élégantes.

si on veut qualifier les 51 personnes d'élégantes. Mais cette règle est une invention machiste en correspondance avec la pensée de l'époque où la règle a été prescrite : « [La hiérarchie entre le masculin et le féminin] remonte au XVIIe siècle lorsqu'en 1647 le célèbre grammairien Vaugelas déclare que « la forme masculine a prépondérance sur le féminin, parce que plus noble ». Dorénavant, il faudra écrire : « Les légumes et les fleurs sont frais » et faire en sorte que l'adjectif s'accorde au masculin, contrairement à l'usage de l'époque qui l'aurait accordé au féminin. En effet, au Moyen Âge, on pouvait écrire correctement, comme Racine au XVIIe siècle : « Ces trois jours et ces trois nuits entières » - l'adjectif « entières » renvoyant alors à « nuits » autant qu'à « jours ». » (Agnès Callamard, « Droits



de l'homme » ou « droits humains »?, Le sexisme à fleur de mots, Le monde diplomatique, mars 1998, page 28)



Exercices

Exercice 1 Chercher des couples de paraphrases dont les verbes principaux correspondent à des sens différents (comme dans l'exemple de la section 2.6 où les verbes principaux de notre couple de paraphrase correspondent au sens 'durer' vs. 'malade'). On pourra ensuite proposer la structure prédicative commune à ce couple de paraphrase, puis les structures syntaxiques profondes des deux phrases proposées.

Exercice 2 Quels sont les quatre moyens d'expression du langage? (Le premier étant les mots.)

Exercice 3 Expliquez en quoi la phrase « Excellent, ce café! » dévie des règles générales du français et proposez une description.

Quel problème pose la traduction de son sac en anglais? Qu'estce que cela illustre comme différence grammaticale entre les deux langues?

Exercice 5 Pour vous rendre compte de la complexité du choix entre l'article défini et l'article indéfini, essayez d'expliquer (de manière claire et compréhensible pour un apprenant du français, japonais par exemple) quand il faut dire « *Où puis-je trouver les toilettes*? » et quand faut-il choisir l'article indéfini « Où puis-je trouver des toilettes? ».

Exercice 6 Il ne faut pas croire que le système de politesse du français est beaucoup plus simple que celui du japonais. Essayez d'expliquer (de manière claire et compréhensible pour un apprenant du français par exemple)

- quand et avec qui on peut utiliser les verbes bouffer, manger ou dîner pour parler d'une personne en train de prendre son repas du soir?
- quelles sont les règles (tutoiement, utilisation du nom/prénom) dans la situation suivante : un professeur tutoie normalement sa collègue, Mme Marie Dupont. Maintenant, il la présente et s'adresse à elle devant un groupe d'étudiants.



Lectures additionnelles

Si l'on est intéressé par l'inventaire des significations morphologiques que les langues nous obligent à communiquer, on pourra consulter le deuxième volume du monumental cours de morphologie d'Igor Mel'čuk. La citation de Jakobson de la section ?? est extraite d'un article très lisible. On peut aussi lire le très bon article publié dans le New York Times Magazine par Guy Deutscher. L'article d'Edward Sapir sur le nootka se trouve dans sa sélection d'écrits. Les autres sujets abordés dans ce chapitre seront largement développés dans la suite de l'ouvrage.

Franz Boas1938 Language, General Anthropology, Boston.

Roman **Jakobson1959** On linguistic aspects of translation, *On translation*, *3*, 30-39.

Guy Deutscher, Does your language shape how you think?, article du *New York Times Magazine* du 26 aout 2010. (http://www.nytimes.com/2010/08/29/magazine/29language-t.html)

Igor Mel'čuk (??) Cours de morphologie générale, 5 volumes, Presses de l'Université de Montréal/CNRS.

Edward Sapir1973 Selected Writings of Edward Sapir, édites par D. Mandelbaum, University of California Press, Berkeley.





Citations originales

Citations de la section ??.

Jakobson (1959) :

Languages differ essentially in what they must convey and not in what they may convey.

Boas (1938:132):

We have to choose between these aspects, and one or the other must be chosen.



Corrections des exercices

Corrigé 1 Il existe de nombreux exemples. Nous en proposons deux :

- (6) Ali pleure parce que Zoé est partie.
- (7) Le départ de Zoé a fait pleurer Ali.
- (8) Les Chinois consomment de plus en plus de viande.
- (9) La consommation de viande des Chinois a augmenté.

Pour la représentation sémantique et les représentations syntaxiques profondes, voir le Chapitre ??. Vous pouvez néanmoins vous essayer dès maintenant à proposer des représentations pour les exemples ci-dessus.

Corrigé 2 Voir la section ??.

Corrigé 3 Nous avons dit que le français demandait que la racine de l'arbre soit un verbe, mais il est également possible que ce soit un adjectif. Dans ce cas, l'argument de l'adjectif ne peut pas être réalisé comme sujet : pour qu'il le soit, il faut ajouter la copule (ce café est excellent), qui devient la tête syntaxique. Il est néanmoins possible d'ajouter l'argument de l'adjectif comme un élément détaché (voir la partie 6 sur la macrosyntaxe).

Corrigé 4 Le français son sac peut être traduit en anglais par his bag ou her bag selon que le propriétaire est un homme ou une femme. Cela illustre le fait que le nom en français a la catégorie du genre et que le déterminant possessif doit s'accorder en genre avec le nom qu'il détermine (son sac vs. sa valise), alors que l'anglais qui n'a pas de genre nominal, distingue pour les référents d'un pronom possessif s'il s'agit d'un non humain (its), d'un humain féminin (her) ou d'un humain masculin (his). Le français n'exprime pas cette information. Une langue peut exprimer les deux informations en même temps, comme l'allemand (sein Koffer, ihr Koffer, seine Tasche, ihre Tasche, où Koffer est un nom masculin qui signifie 'valise', Tasche un nom féminin qui signifie 'sac', sein signifie 'his' et ihr signifie 'her').

Corrigé 5 L'article défini présuppose qu'il existe des toilettes : on l'emploie quand on sait qu'on est dans un endroit où il y a des toilettes. Au contraire, on préfère utiliser l'indéfini quand on n'est pas sûr qu'il y ait des toilettes à proximité ou qu'on pense qu'il peut y en avoir plusieurs.

Corrigé 6 Cette situation est complexe puisqu'on s'adresse simultanément à Marie Dupont et aux étudiants, ce qui peut nous amener à vouvoyer Marie Dupont, alors qu'on la tutoie dans d'autres conditions.

3 La modélisation : Préciser l'objectif de notre étude

3.1 Définition

Ce chapitre tente de caractériser ce qu'est un modèle linguistique et précise le type de modélisation dans lequel s'inscrit notre présentation de la syntaxe. Il n'est pas essentiel à la lecture de la suite de l'ouvrage, mais permet de lever certains présupposés méthodologiques.

Nous appelons modèle un objet construit par le scientifique afin de simuler les propriétés de l'objet d'étude. L'objet d'étude est déterminé par le CADRE THÉORIQUE. Notre position théorique est de considérer que nos objets d'étude sont les langues vues comme des correspondances entre sens et textes. Ce que nous entendons par sens et textes est également déterminé par le cadre théorique (voir les sections 1.3 sur *Sons et textes* et 1.5 sur *Sens et intention communicative*). Le sens est représenté dans le modèle par un objet du modèle que nous appelons la REPRÉSENTATION SÉMANTIQUE. Un modèle, dans ce cadre théorique, devra donc être capable de construire pour chaque représentation sémantique tous les textes correspondants et pour chaque texte toutes les représentations sémantiques correspondantes.

Un modèle permet de faire des prédictions : par exemple, pour un sens, ou plus exactement pour une représentation sémantique, qu'on n'a jamais envisagé avant, on doit être capable de construire les textes qui lui correspondent. Il en découle qu'un modèle est falsifiable : on peut évaluer si les prédictions du modèle correspondent à notre cadre théorique et dire par exemple si tel texte ne peut pas correspondre à tel sens et si un texte donné ne peut correspondre à aucun sens et n'est donc pas un texte de la langue. On peut ainsi construire un contre-exemple, c'est-à-dire une association sens-texte qui n'est pas prédit par le modèle, mais qui appartient à notre objet d'étude, ou, inversement, une association prédite par le modèle, mais qui n'appartient pas à la langue (telle que nous l'envisageons dans notre cadre théorique). L'adéquation du modèle aux données est sa capacité à faire de bonnes prédictions.

Il faut distinguer la falsifiabilité d'un modèle et la RÉFUTABILITÉ de la théo-

rie dans laquelle se situe ce modèle. Il relève des choix théoriques de prendre en compte ou pas tel ou tel phénomène, comme par exemple les limitations mémorielles ou les erreurs de performance (voir la section ?? sur Langue et parole, compétence et performance). On peut réfuter le choix théorique de prendre ou non en compte de tels phénomènes. Mais une fois le choix théorique fait, le modèle doit s'y conformer. Évidemment, la frontière entre théorie et modèle est mouvante et il est tentant d'adapter la théorie aux résultats du modèle. (Voir compléments dans le corrigé de l'exercice 2.)

La PORTÉE du modèle mesure l'ambition théorique du modèle. Par exemple un modèle qui ne prend pas en compte la sémantique aura une portée moins grande qu'un modèle proposant une représentation du sens. La portée peut également concerner les données : un modèle qui prend en compte les erreurs de performance aura une portée plus grande qu'un modèle qui se limite aux productions idéales du locuteur.

La COUVERTURE du modèle est l'ensemble des données de l'objet d'étude qui ont été décrite par le modèle. Plus sa couverture est grande, meilleur est le modèle.

L'ÉCONOMIE du modèle mesure la quantité de paramètres nécessaires à la modélisation d'un phénomène. Plus un modèle est économique, meilleur il est. Ceci repose sur l'idée qu'il y a une certaine économie dans les organismes naturels et que par le biais de la sélection naturelle les systèmes les moins économiques tendent à être éliminés.

La FLEXIBILITÉ du modèle est sa capacité à être adapté à un grand nombre d'objets par la modification d'un minimum de paramètres. Dans le cas des langues naturelles, il s'agit de pouvoir rendre compte des variations dialectales entre locuteurs, ainsi que des variations entre les différents états de langue d'un même locuteur au cours de l'apprentissage.

3.2 Modèle d'une langue ou modèle de la langue



Chaque langue du monde est un objet d'étude et notre objectif premier est de construire des modèles pour chaque langue. Il y a néanmoins quelque chose de commun dans le fonctionnement des différentes langues et l'on peut, lorsqu'on a étudié suffisamment de langues particulières, être tenté d'extraire ce noyau commun que l'on peut appeler la langue ou la faculté de langage. Les propriétés communes à toutes langues sont dites universelles.

Certains linguistes, comme Noam Chomsky, pensent qu'il existe une grammaire commune à toutes les langues qu'on appelle la GRAMMAIRE UNIVERSELLE



et que cette grammaire est innée. Il est en effet assez légitime de penser que, de même que nos mains et nos pieds se sont spécialisés pour réaliser des tâches particulières, une partie de notre cerveau, qui est l'organe du langage, doit s'être spécialisé pour cette tâche. On va alors chercher dans l'élaboration du modèle à bien distinguer PRINCIPES et PARAMÈTRES: les principes sont les universaux innés de la langue que chaque être humain va paramétrer lors de l'acquisition de sa langue. Ces paramètres constituent les idiosyncrasies de chaque langue. Le terme vient du grec idios 'propre' et synkrasis 'constitution'. En médecine ou en psychologie, le terme réfère souvent à un comportement propre à une personne, une réaction qui diffère de personne à personne. En linguistique, on s'intéresse peu aux spécificités langagières d'une personne, mais on s'intéresse aux spécificités d'une langue par rapport à la (faculté humaine de) langue. L'opposition entre principes et paramètres évoque une machine générique qui n'est pas construite différemment pour chaque utilisateur, mais qui permet à l'utilisateur quelques ajustements personnels. Comme pour une machine, ces ajustements personnels - les paramètres - ne sont pas libres et sont interdépendants : on ne peut choisir que certaines valeurs et le choix d'une valeur peut limiter les choix dont on dispose pour une autre valeur.

Une modélisation d'une langue est meilleure quand elle permet facilement le paramétrage d'un sous-langage, d'un dialecte ou d'un registre de langue : on aspire à une analyse du français où un petit changement dans les paramètres - un « paramétrage » - nous donne les grammaires du français écrit journalistique, oral de conversation, de Marseille, de banlieue parisienne et même les différences d'acceptation individuelle. Un tel modèle est meilleur qu'un modèle où ces variantes du français nécessitent des descriptions tout à fait distinctes.

Exemples de propriétés universelles de la langue et de spécificités du français

- Chaque langue parlée a un système phonologique avec un nombre fini de phonèmes. Mais le français a des voyelles nasales (en, on, in) que la majorité des langues ne possèdent pas.
- Toutes les langues ont des syllabes, c'est-à-dire que les locuteurs coarticulent certains sons. Le français permet l'enchaînement de trois consonnes, comme à l'initiale du mot structure, mais d'autres langues ne le permettent pas, comme par exemple le japonais ou le yoruba (une langue très diffusée en Afrique de l'Ouest).
- Toutes les langues ont une structure syntaxique hiérarchique. Le français possède des mots, les pronoms relatifs, qui peuvent à la fois marquer la subordination et jouer un rôle dans la subordonnée (par exemple dans le livre que Pierre lit, que est à la fois un subordonnant et le complément d'objet

direct du verbe LIRE), mais la majorité des langues n'ont aucun mot comme ça.

- Si une langue distingue singulier et pluriel pour les objets non animés, alors, il faut aussi faire cette distinction pour les objets animés, alors que l'inverse n'est pas vrai. Il existe des langues qui possèdent un pluriel grammatical seulement pour les objets animés, par exemple le japonais.
- Le fait de posséder un système flexionnel, comme la conjugaison des verbes en français ou la déclinaisons des noms en latin, n'est pas universel. En effet, le chinois n'a de flexion pour aucune classe de mots.

Quels paramétrages sont possibles? Beaucoup de langues, mais pas toutes, obligent le locuteur à indiquer s'il parle d'un seul ou de plusieurs objets, mais aucune langue n'oblige le locuteur à indiquer la couleur de l'objet décrit. Dans aucune langue, le sens d'un mot dépend de la position de ce mot dans la phrase (début, 2^e place, 3^e place, etc.) et il est probable que les humains sont incapables d'apprendre une telle langue pourtant imaginable (en français, on a quelque chose qui s'apparente à ça avec la différence d'interprétation qu'il y a pour certains adjectifs entre la position avant et après le nom : *un grand savant* vs *un savant grand, un jeune marié* vs *un marié jeune*). Un enfant qui apprend à parler n'est donc pas obligé de peser toutes les possibilités théoriques qu'offre la communication dans l'absolu, mais il doit seulement « paramétrer » les choix qui lui sont offerts. La question de savoir quelles sont les contraintes innées qui président à ces choix reste un important sujet de recherche en linguistique.

3.3 Modélisation et théorie

Pour mieux comprendre ce qu'est la modélisation, nous allons faire un parallèle avec la physique.

3.3.1 Un exemple de modélisation en physique

Tout le monde peut observer les marées au bord de l'océan et les décrire. On verra ainsi que les marées respectent un cycle régulier d'un peu plus de 12 heures entre deux marées hautes, ainsi qu'un cycle d'environ 14 jours entre deux grandes marées et encore un cycle annuel pour l'amplitude des grandes marées, le sommet étant atteint pour les marées d'équinoxe deux fois par an. On pourra ainsi relever très précisément chaque jour l'heure des marées et leur hauteur (leur coefficient) et noter que ces hauteurs varient selon une courbe plus ou moins sinusoïdale. Jusque-là, on a décrit le phénomène des marées, mais on ne l'a pas

modélisé. Si la description est suffisamment poussée et qu'on dispose de beaucoup de données, on pourra, par des méthodes statistiques, **prédire** le moment et la hauteur des marées suivantes.

On peut pousser la description jusqu'à noter une certaine corrélation entre les mouvements de la mer et les mouvements respectifs de la terre, de la lune et du soleil (la terre fait un tour sur elle-même en 24 h, la lune fait le tour de la terre en 28 jours et la terre fait le tour du soleil en un an). Dès que l'on applique la théorie de la gravitation de Newton et que l'on considère que la lune comme le soleil attirent suffisamment les mers pour les faire bouger par rapport à la croûte terrestre, on obtient un nouveau modèle des marées. Ce modèle permet non seulement de prédire les dates et hauteurs des marées des années à l'avance, mais il permet aussi d'expliquer le phénomène des marées et de comprendre la superposition des trois rythmes sinusoïdaux.

3.3.2 La modélisation de la langue

Considérons maintenant la langue. Tout le monde peut observer des productions langagières et les décrire. On peut compter les occurrences de chaque mot, regarder dans quels contextes elles apparaissent, avec quels autres mots avant et après, etc. On peut même noter des corrélations complexes comme la forme du verbe et la présence de tel ou tel pronom à tel endroit. Si on a pris soin de noter les circonstances dans lequel le texte a été énoncé, on peut pousser la description jusqu'à faire des corrélations entre le contexte d'énonciation et le texte produit. On obtient ainsi une DESCRIPTION qui, combinée avec un modèle statistique, peut avoir une bonne VALEUR PRÉDICTIVE. Certains parlent de MODÈLES DESCRIPTIFS et de MODÈLES PRÉDICTIFS, mais, de notre point de vue, la modélisation commence réellement lorsqu'on se place dans un certain cadre théorique et que l'on fait des hypothèses sur le fonctionnement de la langue. On parle alors vraiment de MODÈLE (THÉORIQUE). Celui-ci sera d'autant meilleur qu'il aura une VALEUR EXPLICATIVE, c'est-à-dire qu'il nous permettra de comprendre non seulement comment nous construisons nos énoncés, mais aussi pourquoi nous devons respecter ce type de contraintes, comment nous apprenons à parler, pourquoi nous faisons tel lapsus, etc. On obtient ainsi un MODÈLE EXPLICATIF.

3.3.3 Du modèle à la théorie

On peut pousser encore plus loin la comparaison entre physique et linguistique. Le phénomène des marées n'est pas seulement une application de la théorie de la gravitation. L'histoire de la pomme de Newton est plaisante, mais il est clair qu'elle n'a été qu'un déclencheur. Newton ne cherchait pas à résoudre le problème de la chute des pommes, mais souhaitait surtout comprendre l'origine du mouvement des planètes et le phénomène des marées et c'est l'observation de ces phénomènes qui lui a permis d'élaborer et de valider la théorie de la gravitation. Dès qu'on fait l'hypothèse que les masses s'attirent, tous ces problèmes — la pomme, les planètes et les marées — trouvent une solution simple. C'est ce qu'on appelle un raisonnement par abduction, où on pose une hypothèse A, car elle est l'explication la plus simple à une observation C. C'est par l'observation des marées et de la chute des pommes qu'on peut émettre l'hypothèse que les masses s'attirent.

Il en va exactement de même pour la théorie linguistique : ce sont tous les phénomènes observés dans les langues qui nous permettent d'élaborer, par abduction, une théorie linguistique. C'est de cette façon que, dans la section 2 *Produire un énoncé*, nous avons été amenés à faire l'hypothèse qu'il existe une structure syntaxique hiérarchique sous-jacente aux énoncés, sur laquelle s'appuient les contraintes linguistiques. Nous allons illustrer à nouveau ce point dans l'encadré qui suit sur les phénomènes d'accord.

3.4 Un exemple − l'accord − de la description à l'explication **②**

Nous allons illustrer notre propos précédent par un exemple, celui des règles d'accord.

Étape 1. Description de la concordance des formes du nom et du verbe. On peut remarquer, en français, que la forme du verbe varie en fonction du sujet : Marie dort vs Marie et Pierre dorment, Marie finit de manger vs Marie et Pierre finissent de manger, etc. On poussera la description jusqu'à noter que les segments qui peuvent aller dans l'environnement ___ dort et ___ finit de manger sont les mêmes (Marie, elle, mon amie, la dame, etc.) et donner un nom à cette classe : les groupes nominaux singuliers.

Étape 2. Énoncer la règle d'accord. Voici la règle : en français, il y a deux nombres — singulier et pluriel — et le verbe conjugué s'accorde en nombre avec son sujet. Cette règle paraît élémentaire (il y en a des plus complexes comme la règle d'accord du participe passé en français), mais elle ne l'est pas tant, car elle suppose que l'on sait définir le sujet d'un verbe et le reconnaître. L'histoire de la linguistique montre que ce ne fut pas chose facile et que définir correctement la notion de sujet, c'est déjà élaborer un début de théorie syntaxique. Modéliser cette règle par exemple pour implémenter un correcteur automatique capable



d'assurer l'accord du verbe avec le sujet est encore une autre affaire. Aujourd'hui les correcteurs grammaticaux ne sont généralement pas capables de retrouver les sujets des deux verbes de la phrase « La pièce dans laquelle veulent jouer les enfants est trop petite » (et ils le seraient encore moins s'il y avait une faute d'orthographe).

Étape 3. Pourquoi il y a des règles d'accord dans les langues. Pour bien modéliser la règle d'accord, il faut comprendre quel est son rôle dans le système. Les règles d'accord servent à marquer les relations entre les mots de manière à aider le destinataire à reconstruire la structure de l'énoncé et son sens. Il y a plusieurs manières de marquer ces relations :

- 1. Une marque qui dépend de la nature de la relation : c'est la rection. Il peut s'agir d'un CAS, c'est-à-dire d'une marque sur le dépendant, comme en latin (Petrus 'Pierre' → Petri cani 'le chien de Pierre'), d'une préposition, c'est-àdire d'un mot, comme en français (le chien de Pierre) ou d'une marque sur le gouverneur comme en wolof (xaj bi 'le chien', lit. chien le \rightarrow xaj**u** Peer bi 'le chien de Pierre').
- 2. Une marque d'accord : cette marque peut se trouver sur le dépendant et reprendre une caractéristique du gouverneur (accord de l'adjectif avec le nom) ou l'inverse (accord du verbe avec son sujet).
- 3. Un ordre fixe : la position du dépendant par rapport au gouverneur est très contrainte. Par exemple, en anglais, un adjectif précède le nom dont il dépend et l'objet direct suit le verbe dont il dépend.
- 4. La prosodie : dans la phrase ambiguë Pierre regarde la fille avec un télescope, le groupe prépositionnel avec un télescope peut dépendre de regarde ou de la fille. Un contour prosodique regroupant la fille avec un télescope désambiguïsera la phrase.

Les différentes techniques peuvent se combiner, comme en français où le sujet possède une position contrainte (devant le verbe en général), déclenche un accord du verbe et varie en cas lorsqu'il s'agit d'un pronom (il dort, Marie le regarde, Marie lui parle).

Présenter l'accord comme nous venons de le faire, c'est prendre une position théorique : nous considérons que les mots se connectent entre eux et forment une structure hiérarchique et que les phénomènes d'accord en découlent normalement. De la même façon, la théorie de la gravitation fait l'hypothèse que les masses s'attirent et montre que les marées découlent naturellement de cette hypothèse.

3.5 Modèle déclaratif

Comme nous l'avons dit, un modèle linguistique doit être capable d'associer une représentation sémantique à des textes et vice versa. On distingue dans le modèle l'ensemble des **connaissances** nécessaires pour effectuer cette association de la procédure qui permet d'activer ces connaissances et de réaliser l'association. Un modèle qui sépare connaissances et procédure est dit déclaratif. Sinon le modèle est dit procédural. On appelle généralement grammaire formelle un modèle déclaratif d'une langue. Le terme *grammaire* employé comme ceci inclus la description du lexique de la langue.

Supposer que les langues possèdent des modèles déclaratifs est une hypothèse forte et difficile à vérifier. Elle repose sur l'idée que les connaissances qui permettent de parler une langue et de la comprendre sont peu ou prou les mêmes. Une grammaire commune à la production et à l'analyse est dite RÉVERSIBLE. On aura ensuite des procédures distinctes pour la production et l'analyse. La procédure d'analyse est plus tolérante, puisque les locuteurs comprennent des énoncés qu'ils ne seraient pas capables de produire; les contraintes de la grammaire devront donc être relâchées en analyse.

Un modèle déclaratif est normalement un Modèle de la compétence. En effet, les erreurs de performance doivent être imputées à des problèmes rencontrés lors de l'activation des connaissances. Néanmoins opter pour un modèle déclaratif et une séparation entre connaissances et procédures ne signifie pas que nous rejetons la procédure hors du modèle. Notre modèle de la compétence doit être inclus dans un modèle complet de la langue prenant en compte la mise en œuvre du modèle déclaratif. Ce modèle complet est un Modèle de la Performance.

3.6 Modèle génératif, équatif et transductif 🕏

Noam Chomsky a révolutionné la linguistique en 1957 dans son ouvrage *Syntactic structures* en définissant son objet étude comme un problème mathématique. Il pose qu'une langue est l'ensemble potentiellement infini des phrases grammaticales de cette langue et qu'une grammaire est un système mathématique comportant un nombre fini de règles capable de générer l'ensemble des phrases grammaticales d'une langue. Ce courant sera appelé la GRAMMAIRE GÉNÉRATIVE. Plus tard, Chomsky renoncera à la présentation générative de la langue, mais on continuera d'appeler son école de pensée la grammaire générative, ce qui prête souvent à confusion. La naissance de la grammaire générative a coïncidé avec la naissance de la cybernétique et de l'informatique théorique et les deux



mouvements se sont fortement influencés, langues naturelles et langages de programmation ayant été vus comme des objets de natures similaires.

Dans sa version initiale, la grammaire générative ne traite pas le sens, considéré comme difficilement accessible. Les phrases, notamment dans les versions formalisées du modèle, sont traitées comme de simples suites de caractères, c'està-dire, pour reprendre notre terminologie, uniquement des textes. La grammaire ne se fixe alors comme principal objectif que de générer les textes d'une langue. En fait les grammaires proposées par Chomsky offrent une analyse syntaxique des phrases et définissent donc, indirectement, une correspondance entre textes et représentations syntaxiques. D'autres chercheurs ont proposé, comme Aravind Joshi en 1975, des GRAMMAIRES D'ARBRE permettant de générer simultanément une phrase et sa représentation syntaxique, construisant ainsi les premières grammaires de correspondance génératives.

Une GRAMMAIRE DE CORRESPONDANCE est une grammaire définissant la correspondance entre deux ensembles de structures, par exemple des représentations sémantiques et des textes (voir l'encadré ?? sur La Théorie Sens-Texte). On peut utiliser trois types de procédures pour définir une correspondance avec une grammaire de correspondance : procédure générative, équative ou transductive. Une PROCÉDURE GÉNÉRATIVE est une procédure qui va générer la correspondance, c'est-à-dire l'ensemble des couples en correspondance; au lieu de générer uniquement un texte, la grammaire génère simultanément le texte et son sens. Une PROCÉDURE ÉQUATIVE est une procédure qui va vérifier pour chaque couple de structures qu'on lui proposera si ces structures se correspondent; cela suppose qu'on fournisse un texte et un sens et la procédure permettra de vérifier que ce texte et ce sens peuvent être associés par la grammaire. Une procédure trans-DUCTIVE est une procédure qui à chaque fois qu'on lui propose une structure est capable de construire toutes les structures qui lui correspondent; dans ce cas, on fournit soit un texte, soit un sens et la procédure construit les sens correspondant au texte ou les textes correspondant au sens. Dans le Chapitre 2, nous avons adopté une procédure transductive pour associer un graphe sémantique à des arbres syntaxiques.

Ces trois types de procédure peuvent être utilisés pour présenter une même grammaire. Ce qui distingue ces trois procédures est le nombre de structures au départ : 0, 1 ou 2. Dans tous les cas, on a un couple de structures à l'arrivée. Dans la procédure générative, on part de rien et on génère simultanément les deux structures en correspondance. Dans les procédures transductives, on a une structure au départ et on produit l'autre. Dans la procédure équative, on a les deux structures dès le départ et on vérifie qu'elles se correspondent. Nous sché-

matisons ci-dessous les trois procédures. Supposons qu'on veuille associer des graphes sémantiques, représentés par des \boxtimes , à des arbres syntaxiques, représentés par des \triangle . Ce qui distingue les trois procédures, ce sont les structures données au départ : nous les schématisons en noir, tandis que les structures à construire par la procédure sont en blanc :

- (1) procédure générative : 🏻 ⇔ △
- (2) procédures transductives : $\star \Rightarrow \triangle$ ou $\boxtimes \Leftarrow \boxtimes$
- (3) procédure équative : ★ ⇔ 🛭

Pour des raisons historiques que nous venons de rappeler, la procédure générative est souvent privilégiée. La procédure équative est généralement la procédure la plus élégante pour présenter un modèle déclaratif et elle tend à se généraliser sous le nom de GRAMMAIRES DE CONTRAINTES. Mais des trois procédures, c'est la procédure transductive qui est descriptivement la plus pertinente, car c'est ce type de procédure que les locuteurs utilisent quand ils parlent : ils doivent à partir d'un sens lui faire correspondre un texte et, à l'inverse, quand ils écoutent quelqu'un qui parle, ils partent d'un texte et doivent lui donner un sens.

3.7 Modèle symbolique

Une des propriétés remarquables des langues est l'utilisation d'un petit nombre de sons — les phonèmes — pour construire les signifiants de tous les éléments lexicaux de la langue. Cela met en évidence notre capacité à catégoriser, c'est-à-dire à identifier, dans la multitude de signaux de parole auxquels nous sommes confrontés lors de l'apprentissage de notre langue, un nombre fini de symboles, c'est-à-dire d'éléments qui ont une portée symbolique et qui possède une valeur d'interprétation. Plus généralement, on peut penser que les fonctions supérieures du cerveau, celles qui sont liées à la cognition et à la pensée consciente, se caractérisent par la capacité à catégoriser et à manipuler des symboles.

On appelle MODÈLE SYMBOLIQUE OU MODÈLE DISCRET OU encore MODÈLE ALGÉBRIQUE un système basé uniquement sur la manipulation algébrique d'un nombre fini de symboles. Par manipulation algébrique, on entend des opérations mathématiques permettant de combiner des configurations de symboles pour créer de nouvelles configurations (voir encadré ci-dessous sur le calcul symbolique). Un modèle qui ne « discrétise » pas est dit CONTINU. La question se pose de savoir si un modèle linguistique doit ou non être symbolique. Les arguments contre les modèles symboliques sont assez nombreux :



- l'acceptabilité des énoncés n'est pas binaire : il semble y avoir un continuum entre les énoncés acceptables et les énoncés inacceptables;
- les catégories syntaxiques sont assez floues; on trouve de nombreux éléments à la frontière de plusieurs catégories;
- les unités lexicales sont généralement polysémiques et il est difficile de déterminer combien de sens peut avoir exactement une unité lexicale;
- la prosodie, qui joue un rôle non négligeable dans l'expression du langage, semble difficilement catégorisable (même si la ponctuation est une forme de catégorisation de certains contours prosodiques).

Malgré cela, on arrive à fournir des modèles symboliques des langues assez satisfaisants et la plupart des modèles théoriques sont basés sur des règles manipulant des symboles.

On peut rendre compte des différents niveaux d'acceptabilité en modifiant un peu un modèle symbolique. Deux directions au moins ont été envisagées. La première, exploitée par la Théorie de l'Optimalité de Alan Prince et Paul Smolensky1993, consiste à traiter les règles comme des contraintes éventuellement violables. De plus, les contraintes peuvent être rangées par ordre d'importance. Plus un énoncé viole de contraintes et plus ces contraintes sont importantes, moins il est acceptable.

L'autre direction consiste à pondérer les règles. On peut alors associer un score à chaque énoncé en fonction des règles qui ont permis de le produire. Cette technique est surtout utilisée en analyse pour désambiguïser : lorsque plusieurs analyses sont possibles pour un énoncé, on privilégie celle qui a le meilleur score. Des grammaires de ce type sont généralement construites en pondérant les règles selon leur probabilité d'apparition dans un corpus syntaxiquement annoté calculée à partir d'une analyse statistique de leurs occurrences. On obtient ainsi un mo-DÈLE STOCHASTIQUE, dont la base reste un modèle symbolique.

Nous montrerons dans cet ouvrage les différents problèmes que pose l'identification des unités d'une langue et leur catégorisation.

3.8 Calcul symbolique et grammaires catégorielles 🕏



Le premier linguiste mathématicien montrant qu'on pouvait vérifier la bonne formation d'une phrase par un CALCUL SYMBOLIQUE est probablement le polonais Kazimierz Ajduckiewicz en 1935. Voici son idée. Pour montrer que « Pierre dort » est une phrase nous allons associer à chaque mot une catégorie complexe :

- Pierre forme à lui seul un groupe nominal, nous lui associons la valeur GN;

 dort peut former une phrase P à condition qu'on le combine avec un GN : nous lui associons la valeur.

Nous pouvons maintenant calculer la valeur associée à Pierre dort :

(4) Pierre *dort*
$$GN \cdot = P$$

Cette valeur est calculée en combinant les catégories associées à chaque mot et en simplifiant comme on le fait avec des fractions ordinaires. L'analyse d'une phrase devient un CALCUL ALGÉBRIQUE similaire au calcul numérique $(a. \frac{b}{a} = b)$.

Nous avons construit un début de grammaire capable de vérifier que chaque verbe conjugué a un sujet. Une grammaire associant ainsi des catégories complexes à chaque mot est appelée une GRAMMAIRE CATÉGORIELLE. Il s'agit du premier exemple de GRAMMAIRE FORMELLE connu.

Ce calcul a été ensuite repris par Yehoshua Bar-Hillel en 1953, qui a montré que si l'on veut modéliser les contraintes d'ordre, il fallait distinguer ce qu'on combine à droite de ce qu'on combine à gauche. On va donc associer à *dort* la catégorie GN\P (qui se lit « GN sous P ») indiquant que le GN avec lequel le verbe doit se combiner pour former une P doit se trouver à gauche. Analysons *Pierre mange une banane* avec cette grammaire. Les mots de cette phrase ont les catégories *Pierre* := GN, *banane* := D\GN, un := D et $mange := (GN\P)/GN$. Le calcul est le suivant :

(5) Pierre mange une banane

```
(6) GN · (GN\P)/GN · D · D\GN
GN · (GN\P)/GN · GN
GN · GN\P
P
```

Cette séquence de mots est bien reconnue comme une phrase par notre grammaire. De plus, la structure du calcul est un arbre que l'on peut interpréter comme la structure syntaxique de la phrase.

Joachim Lambek a montré en 1958 que la règle de combinaison des catégories pouvait être interprétée comme une inférence logique. On peut en effet voir GN\P comme une implication GN \rightarrow P à interpréter comme « si on me donne un GN à gauche, je formerai un P ». La règle de combinaison devient alors : « de GN et de GN \rightarrow P, je déduis P », ce qui n'est autre que le *modus ponens*, la règle de déduction de base de la logique (« de p et de $p \rightarrow q$, je déduis q »). L'analyse d'une

phrase devient maintenant un CALCUL LOGIQUE. Le calcul devient une *preuve* que la suite de mots considérée au départ est bien une phrase. Il est intéressant de remarquer que la logique ainsi construite diffère de la logique classique, puisqu'elle est **sensible aux ressources** (donner deux GN n'est pas équivalent à en donner un seul) **et à l'ordre** (GN \rightarrow P et P \leftarrow GN ne se comportent pas pareil, l'un attend un GN avant et l'autre après). Une telle logique, appelée logique linéaire, a des applications dans des domaines très éloignés de la linguistique et notamment en robotique où les actions doivent être effectuées dans un ordre bien précis.

Dans l'encadré ?? sur la *Grammaire de réécriture*, nous présenterons un autre exemple de grammaire formelle qui a marqué la deuxième moitié du 20^e siècle : les grammaires de réécriture hors-contexte introduites par Noam Chomsky en 1957.

3.9 Modularité et stratification

Un module est un sous-système du modèle suffisamment autonome pour effectuer seul une partie des calculs nécessaires à la production d'un énoncé. Les travaux en neurologie semblent accréditer l'idée que le cerveau possède un fonctionnement modulaire et que, lors de la production d'un énoncé, différentes aires cérébrales sont sollicitées avec des tâches différentes. Néanmoins les connaissances sur l'architecture du cerveau sont encore insuffisantes pour que se dégage une vision claire des différents modules que devrait avoir un modèle linguistique et sur la façon dont ces modules coopèrent.

Dans la suite de cet ouvrage, nous montrerons qu'il existe plusieurs types d'unités linguistiques et plusieurs modes d'organisation de ces unités. Une architecture possible est alors de considérer que chaque niveau d'organisation fournit une strate et que le passage du sens au texte se fait en plusieurs modules qui permettent de passer d'une strate à l'autre, l'un à la suite de l'autre. Un modèle de ce type est dit stratificationnel. La façon la plus simple d'utiliser un modèle stratificationnel est d'avoir une séquence de modules qui fonctionnent à la suite l'un de l'autre : un premier module prend en entrée le sens et fournit au deuxième module une structure complète de la strate suivante et ainsi de suite. Une telle architecture est dite linéaire ou en pipeline. Cette architecture s'oppose à une architecture distribuée où toutes les strates peuvent communiquer. Même dans une architecture linéaire, on peut faire que tous les modules fonctionnent simultanément et que chaque module traite les données que lui fournissent les autres modules au fur et à mesure qu'elles arrivent. Une telle procédure est dite incrémentale. Le modèle que nous défendons est stratifié, modulaire, incrémenter de la suite de l'autre. Une telle procédure est dite incrémenter de la mesure qu'elles arrivent. Une telle procédure est dite incrémenter de la mesure qu'elles arrivent. Une telle procédure est dite incrémenter de la mesure qu'elles arrivent. Une telle procédure est dite incrémenter de la mesure qu'elles arrivent. Une telle procédure est dite incrémenter de la mesure qu'elles arrivent. Une telle procédure est dite incrémenter de la mesure qu'elles arrivent. Une telle procédure est dite incrémenter de la mesure qu'elles arrivent. Une telle procédure est dite incrémenter de la mesure qu'elles arrivent.

tal et en grande partie linéaire. Nous n'avons aujourd'hui aucun moyen de valider ou d'invalider une telle architecture.

3.10 La Théorie Sens-Texte

La Théorie Sens-Texte est la théorie qui a le plus fortement inspiré les auteurs de cet ouvrage. Il s'agit d'un modèle développé autour d'Igor Mel'čuk à partir de 1965, d'abord en Union Soviétique, puis au Canada et en Europe. Ce modèle est fortement stratifié : il suppose l'existence de 5 niveaux de représentation intermédiaires entre le sens et le texte, soit 7 niveaux en tout. Le système est divisé en 6 modules permettant de passer d'un niveau à l'autre. Il s'agit d'une architecture linéaire que nous présentons ci-dessous :

(7) sens = représentation sémantique représentation syntaxique profonde représentation syntaxique de surface représentation morphologique profonde représentation morphologique de surface représentation phonologique profonde texte = représentation phonologique de surface

Chaque ⇔ représente un module effectuant la correspondance entre deux niveaux de représentation adjacents. On trouvera une présentation et une justification de cette architecture dans le cours donné par Igor Mel'čuk en 1997 au collège de France, *Vers une linguistique Sens-Texte*, disponible en ligne.

Dans cet ouvrage, nous présentons les quatre niveaux supérieurs de représentation. Notre présentation peut être assez différente de celles que l'on trouve dans les travaux d'Igor Mel'čuk. Notre objectif est de justifier toutes les structures que nous introduirons, quitte parfois à remettre en question le statut ou la nature des représentations utilisées en Théorie Sens-Texte ou dans d'autres théories comparables, notamment la *Lexical Functional Grammar (LFG)* de Joan Bresnan et Ronald Kaplan, qui possède également une architecture stratificationnelle.

3.11 Modélisation des langues et ordinateur

Le développement « à la main » d'un modèle linguistique est un travail considérable. Un dictionnaire de français courant possède 60 000 mots et l'on évalue à près d'un million le nombre d'unités lexicales si l'on y inclut les expressions figées et que l'on compte les différentes acceptions de chaque lexème. Le

lexique des constructions grammaticales est encore mal connu et certainement sous-évalué. La combinatoire de n'importe quelle description sérieuse d'un phénomène grammatical est tellement importante qu'il est difficile de voir la description dans son ensemble. L'ordinateur est alors le seul moyen pour combler cette difficulté. Les ordinateurs possèdent aujourd'hui des capacités de calcul suffisantes pour tester la plupart des modèles imaginables (et théoriquement raisonnables).

Aujourd'hui les modèles capables d'effectuer des correspondances entre sens et textes sont très fragmentaires. On trouve par exemple des modèles de ce type assez performants, mais limités à la production de bulletins météo et manipulant donc un vocabulaire restreint. Les modèles informatiques couvrants se limitent à la correspondance entre textes et représentations syntaxiques de surface et ils font aujourd'hui pour cette tâche d'analyse superficielle près d'une erreur par phrase en moyenne.

La validation des théories linguistiques n'est pas le seul intérêt de l'implémentation des modèles : le développement de modèles informatiques possède un réel intérêt économique, puisque la langue est au centre de toutes les activités sociales humaines. Le développement de modèles informatiques et de ressources formelles tels que lexiques, grammaires ou corpus annotés s'appelle la linguistique computationnelle est incluse dans le domaine plus vaste du traitement automatique des langues ou TAL, qui s'intéresse également à toutes les applications que l'on peut développer à partir de tels modèles. Les applications industrielles du TAL constituent l'ingénierie linguistique.

Le TAL ne nécessite pas toujours de grandes connaissances en linguistique et certaines méthodes de traitement du langage n'ont pas grand-chose à voir avec la modélisation des langues. Par exemple, pour décider en quelle langue est une page sur la Toile le moyen le plus simple et le plus performant est de faire une statistique des séquences de trois lettres et de comparer avec les statistiques pour les langues que l'on souhaite reconnaître. Nul besoin de lexique et encore moins de connaissances en grammaire. De la même manière, il est imaginable de faire une étude statistique sur les marées des cent dernières années et en déduire de bonnes prédictions sur les marées à venir. Par contre, il semble clair qu'une telle « modélisation » ne nous avance pas dans la compréhension du phénomène des marées et du rôle de la gravitation (voir la section 3.3 sur *Modélisation et théorie*).

Aujourd'hui, beaucoup de modèles en TAL se basent avant tout sur des données statistiques apprises sur de grand corpus, car les résultats sont souvent meilleurs et moins coûteux qu'avec des données entièrement développées à la

main qui s'avèrent généralement incomplètes. La très grande capacité de mémoire de l'ordinateur fait qu'il est souvent plus facile de travailler avec une énorme liste de données (par exemple toutes les constructions possibles pour tous les verbes d'une langue) que de dégager et d'implémenter les régularités derrières (avec leurs exceptions qui existent toujours). Il est possible que les modèles les plus efficaces pour les applications informatiques ne soient pas les modèles les plus satisfaisants du point de vue de leur « explicativité ». À l'heure actuelle, les traducteurs automatiques les plus performants sont ceux qui interrogent de grands corpus bilingues pour rechercher tous les fragments du texte pour lesquels on a déjà une traduction et qui choisissent la plus fréquente (modèle de traduction). Ainsi les traductions sont combinées et lissées à l'aide de corpus unilingue encore plus grands (modèle de langue). De tels systèmes de traduction statistiques obtiennent de meilleurs résultats que les systèmes qui cherchent à simuler un traducteur humain, c'est-à-dire qui cherchent à calculer une représentation sémantique du texte à traduire pour générer à partir de celle-ci un texte dans une autre langue.



Exercices

Exercice 1 Quelle différence de statut faisons-nous entre le sens et la représentation sémantique?

Exercice 2 Quelle est la différence entre falsifiabilité et réfutabilité?

Exercice 3 Quels seraient des modèles respectivement descriptif, prédictif et explicatif du réchauffement climatique?

Exercice 4 Nous avons discuté de la règle d'accord du verbe avec son sujet et du fait que cette règle reposait sur une définition préalable du sujet. Comme définiriez-vous le sujet syntaxique pour le français?

Exercice 5 L'énoncé *La plupart sont verts* remet-il en cause la règle d'accord en nombre du verbe avec son sujet? Comment résoudre le problème?

Exercice 6 Le français possède deux genres (on dit *le soleil* et *la lune* ou *une armoire* et *un tabouret*), l'allemand en possède trois (féminin, masculin et neutre), l'anglais aucun (on a juste un marquage du sexe dans les pronoms), les langues bantoues peuvent avoir jusqu'à une vingtaine de classes nominales différentes. Ce phénomène est facile à décrire, mais qu'est-ce qu'un modèle explicatif pourrait en dire? Pourquoi les langues peuvent avoir ou ne pas avoir des classes d'accord différentes pour les noms?

Exercice 7 Pourquoi le fait qu'on puisse distinguer une question d'une assertion par la seule prosodie (une intonation montante pour « *Tu viens* ? » et descendante pour « *Tu viens*. ») met-il en défaut une architecture linéaire comme celle de la Théorie Sens-Texte ?



Lectures additionnelles

Sur la distinction entre falsifiabilité et réfutabilité, on pourra lire l'article de Lakatos1968 qui fait lui-même référence aux débats entre Popper et Kuhn suite à la réfutation de la théorie de la gravitation de Newton et à sa résolution par la théorie de la relativité d'Einstein.

La distinction entre modèle et théorie est discutée dans Chomsky1957, dont la lecture est incontournable si l'on veut comprendre pourquoi cet ouvrage a marqué un basculement de la linguistique dans le domaine des sciences. Pour Chomsky, une théorie donne un formalisme grammatical

et les modèles sont les grammaires particulières que l'on peut définir avec ce formalisme.

Le raisonnement par abduction a été dégagé par le philosophe américain, Charles S. Peirce (1839-1914). En 1903, il en donne la formulation suivante : « The surprising fact, C, is observe; But if A were true, C would be a matter of course; Hence, there is reason to suspect that A is true. ».

La Théorie Sens-Texte est présentée dans la plupart des ouvrages d'Igor Mel'čuk. Voir les ouvrages dont nous avons déjà parlé dans les trois précédents chapitres.

Noam Chomsky1957 Syntactic structures, MIT Press (traduction française de M. Bradeau, 1969, Structures syntaxiques, Seuil).

Imre Lakatos1968 Criticism and the Methodology of Scientific research Programmes, *Proceedings of the Aristotelian Society*, New Series, vol. 69, Blackwell, pp. 149-186.

Charles S. **Peirce1903** *Harvard lectures on pragmatism*, Collected Papers v. 5, paragraphs 188–189.



Corrections des exercices

Corrigé 1 Tout énoncé linguistique a un sens. Le sens appartient à la langue. La représentation sémantique appartient au modèle de la langue et modélise le sens.

Corrigé 2 La falsifiabilité est une propriété des modèles. Lorsqu'un modèle est faux (c'est-à-dire qu'il fait une mauvaise prédiction), on peut essayer de le réparer en changeant des paramètres. La réfutabilité est une

propriété des théories. Pour réfuter une théorie, il faut montrer que tous les modèles qu'elle propose sont faux, ce qui est très difficile, voire impossible. La réfutabilité d'une théorie se fait donc généralement en proposant une nouvelle théorie dont l'un des modèles prend mieux en compte les données qui posent problème à la théorie précédente.

Corrigé 3 Un modèle descriptif du réchauffement climatique serait une description des relevés de températures en divers point du globe dans les années ou les siècles qui précèdent qui montrerait une augmentation de température. Une analyse statistique des variations de température permettrait de faire des prédictions sur l'évolution de la température dans les années ou siècles à venir (avec l'hypothèse théorique que la température d'une année est corrélée aux températures des années qui précèdent). Un modèle explicatif tenterait de rechercher les causes des variations de température et de corréler ces variations avec un certain nombre de paramètres, comme la consommation d'énergie par les humains. Le modèle permet alors d'affiner la prédiction en fonction de l'évolution de ces paramètres et donc de prédire que la température augmentera encore plus vite si la consommation d'énergie augmente.

Corrigé 4 La notion de sujet syntaxique sera définie dans le Chapitre ??. L'accord du verbe est une des propriétés définitoires du sujet en français, avec la position privilégiée avant le verbe ou l'emploi de pronom comme *il* ou *on*.

Corrigé 5 La plupart est utilisé en français comme pronom pluriel masculin ou féminin. Il s'agit d'une forme figée, d'un sémantème (voir le Chapitre ??), où la n'est plus un marqueur du singulier féminin. Il suffit donc de déclarer la plupart comme un pronom pluriel dont le genre dépend de son antécédent pour assurer l'accord selon les règles habituelles.

Corrigé 6 L'origine des genres en français (et dans les autres langues indo-européennes) est un marquage des sexes pour les noms d'êtres sexués qui s'est propagé à tous les noms par régularisation du système. Pour la plupart des noms, il n'a aucune signification (même si des études montrent

que l'existence des genres a une influence sur la représentation mentale et que les locuteurs d'une langue où *mort* est féminin comme le français personnifieront plus naturellement la mort par une femme que par un homme à l'inverse des locuteurs d'une langue comme l'allemand où *Tod* est masculin). Devoir apprendre des genres pour des noms où cela n'a pas de sens a un coût cognitif. Mais les genres vont permettre de renforcer le marquage des relations syntaxiques au travers des accords en genre. Nous avons vu que l'accord est un des moyens de marquer l'existence d'une relation syntaxique, voire la nature de cette relation (l'accord du verbe avec son sujet permet de caractériser cet élément en tant que sujet).

Corrigé 7 L'interrogation réalisée par la seule prosodie est un exemple d'un sens qui est réalisé directement au niveau phonologique sans aucune incidence sur la syntaxe. La description de cette construction peut être faite par une correspondance directe entre sémantique et phonologie, alors qu'un modèle en pipeline obligerait à passer par la syntaxe et donc à introduire un élément fictif au niveau syntaxique.

Deuxième partie

Les unités de la langue : Les trois composantes du signe linguistique

Présentation

Cette deuxième partie s'intéresse à la délimitation des unités de la langue — les signes linguistiques — selon leur forme, leur combinatoire et leur sens. Le Chapitre 4 présente rapidement pourquoi nous considérons trois types d'unité. Le Chapitre 5 présente l'identification des signes linguistiques et plus particulièrement les unités minimales de forme, les morphèmes, et les contrastent avec les unités minimales pour la combinatoire libre, les syntaxèmes. Le Chapitre 6 présente les unités minimales de sens, les sémantèmes, et les contrastent également avec les syntaxèmes.

4 Trois types d'unités : Morphème, syntaxème, sémantème

4.1 Introduction

« De même que le jeu d'échecs est tout entier dans la combinaison des différentes pièces, de même la langue a le caractère d'un système basé complètement sur l'opposition de ses unités concrètes. On ne peut ni se dispenser de les connaître, ni faire un pas sans recourir à elles; et pourtant leur délimitation est un problème si délicat qu'on se demande si elles sont réellement données. » (Saussure1916 : 149)

Notre objectif est d'étudier la structure de la langue. Nous verrons dans la Partie 3 que, contrairement à ce qu'affirme Saussure dans la citation qui précède, on peut en grande partie s'abstraire de la question des unités lorsqu'on définit la structure syntaxique. Ce qui nous intéresse, ce sont les combinaisons entre les unités et non les unités elles-mêmes. Néanmoins, pour pouvoir parler des combinaisons d'unités, il nous faut dire avant un mot des unités. Nous profiterons de cette discussion sur les unités pour montrer que les unités de la langue, que l'on appelle des SIGNES LINGUISTIQUES, peuvent être appréhendées selon trois points de vue (morphologique, syntaxique et sémantique) et groupées en ensembles de signes différents selon chacun de ces points de vue. Nous nous attarderons en particulier sur la question des unités minimales. Cette question n'est pas fondamentale pour étudier la structure, mais elle n'est pas inutile non plus et elle a l'avantage de pointer les différences entre morphologie, syntaxe et sémantique.

4.2 Double articulation du langage

« L'intention et la capacité de signification [...] sont constitutives du son articulé; et on ne peut rien proposer d'autre pour le distinguer d'une part du cri animal et d'autre part du son musical. » (Humboldt1836 : 60)

Considérons la situation suivante : un acteur vient de faire une performance et vous souhaitez le féliciter, c'est-à-dire lui communiquer le plaisir que vous a

procuré sa performance. Vous avez plusieurs moyens à votre disposition et notamment applaudir ou crier « Bravo! ». Chacune de ces deux réalisations possède une signification du type 'je te félicite', exprimant que celui qui en fait la réalisation souhaite féliciter celui à qui il l'adresse. Chacune des deux réalisations possède une forme spécifique : l'applaudissement se réalise en frappant les deux mains à plat l'une sur l'autre et « Bravo! » se réalise en produisant une séquence sonore particulière. Cette association entre une forme et une signification est appelée un signe. Ces deux actes de communication sont des signes conven-TIONNELS: la forme de ce signe obéit à une convention que s'est fixée un groupe particulier de personnes et qui leur sert à communiquer. De plus, la relation entre leur signification et leur forme est ARBITRAIRE: on pourrait tout aussi bien féliciter quelqu'un en levant les bras et en agitant les mains comme le font les sourds en langue des signes, ou bien prononcer une autre séquence sonore, comme le font par exemple les Chinois qui crient « Hao! » pour communiquer leur appréciation. Inversement l'applaudissement et « Bravo! » pourraient tout aussi bien signifier autre chose que 'je te félicite'.

Au delà de leurs points communs, ces deux signes présentent des différences importantes qui tiennent à la double articulation du langage (voir encadré ?? sur l'origine du terme). La première articulation s'illustre par le fait que le signe bravo peut être combiné à d'autres signes pour former des énoncés plus complexes : « Bravo pour ton excellente prestation! », « Alors là, excuse-moi, mais je ne te dis pas bravo pour ce que tu viens de faire. ». En prenant la question à l'envers, on peut remarquer que la plupart des énoncés de la langue peuvent être décomposés en signes plus élémentaires et que la signification de ces énoncés est la combinaison des significations des signes élémentaires qui les composent. Rien de tel avec l'applaudissement qui ne peut pas être combiné avec un autre signe du même type pour exprimer une signification nouvelle.

La DEUXIÈME ARTICULATION tient au fait que la substance sonore des signes d'une langue peut être décomposée en la combinaison d'un tout petit nombre de sons élémentaires. Dans le signifiant de *bravo*, tout locuteur du français reconnaît cinq sons et chacun de ces cinq sons peut être permuté avec un autre son du français pour donner des mots (potentiels) différents du français : *travo*, *blavo*, *brivo*, *braco*, *brava*. Dans toutes les langues du monde, il existe un ensemble fini de quelques dizaines de sons élémentaires qui permettent de construire les signifiants de tous les énoncés de cette langue.

Les unités de première articulation et leurs combinaisons sont les SIGNES LINGUISTIQUES. Les unités de deuxième articulation, les **segments sonores minimaux**, sont les PHONÈMES. L'étude des phonèmes en tant que telle est en dehors



du champ de cet ouvrage, bien que le système phonologique d'une langue présente des similitudes structurelles avec le système des signes de la langue et que les outils d'investigation des deux systèmes soient en partie similaires.

4.3 Découverte de la double articulation





L'histoire de la mise en évidence de la double articulation du langage est celle de l'invention de l'écriture. Les étapes de l'invention de l'écriture peuvent être tracées ainsi : d'abord les hommes ont créé des PICTOGRAMMES isolés symbolisant par exemple la fécondité ou la chasse, puis des IDÉOGRAMMES associés aux signes linguistiques et combinés pour faire des textes (comme les hiéroglyphes de l'égyptien ancien ou les sinogrammes), puis une ÉCRITURE SYLLABAIRE avec un symbole par syllabe, et enfin un ÉCRITURE ALPHABÉTIQUE avec optimalement une lettre par son (comme en phénicien, puis en grec ancien). On peut dire avec certitude que les savants qui ont élaboré les premières écritures alphabétiques il y a plus de 6000 ans avaient compris la double articulation du langage.

La double articulation du langage a été introduite par Ferdinand de Saussure (Saussure1916 : 26) :

« En latin articulus signifie « membre, partie, subdivision dans une suite de choses »; en matière de langage, l'articulation peut désigner ou bien la subdivision de la chaîne parlée en syllabes, ou bien la subdivision de la chaîne des significations en unités significatives. »

On doit le terme de DOUBLE ARTICULATION à André Martinet1960 :

« La première articulation du langage est celle selon laquelle tout fait d'expérience à transmettre, tout besoin qu'on désire faire connaître à autrui s'analyse en une suite d'unités douées chacune d'une forme vocale et d'un sens. Si je souffre de douleurs à la tête, je puis manifester la chose par des cris. Ceux-ci peuvent être involontaires; dans ce cas ils relèvent de la physiologie. Ils peuvent être plus ou moins voulus et destinés à faire connaître mes souffrances à mon entourage. Mais cela ne suffit pas à en faire une communication linguistique. Chaque cri est inanalysable et correspond à l'ensemble, inanalysé, de la sensation douloureuse. Tout autre est la situation si je prononce la phrase j'ai mal à la tête. Ici, il n'est aucune des six unités successives j', ai, mal, à, la, tête qui corresponde à ce que ma douleur a de spécifique. Chacune d'entre elle peut se retrouver dans de tout autres contextes pour communiquer d'autres faits d'expérience : mal, par exemple, dans il fait le mal, et tête dans il s'est mis à leur tête. [...] Chacune de ces unités de première articulation présente, nous l'avons vu, un sens et une forme vocale (ou phonique). Elle ne saurait être analysée en unités successives plus petites douées de sens. L'ensemble tête veut dire 'tête' et l'on ne peut attribuer à tê- ou à -te des sens distincts dont la somme serait équivalente à 'tête'. Mais la forme vocale est, elle, analysable en une succession d'unités dont chacune contribue à distinguer tête, par exemple, d'autres unités comme bête, tante ou terre. C'est ce qu'on désignera comme la deuxième articulation du langage. Dans le cas de tête, ces unités sont au nombre de trois; nous pouvons les représenter au moyen des lettres t e t, placées par convention entre barres obliques, donc /tet/. »

Si nous reprenons à notre compte la double articulation telle que présentée par Martinet, nous voudrions ajouter une remarque sur ce qui est désigné ici par « cri ». Il est admis aujourd'hui qu'un cri de douleur comme « Aïe! » est bien un mot du français. Un allemand ne dira pas « Aïe! », mais « Au! », (prononcé a-ou) et un anglais « Ouch! » (prononcé a-outch). Ces signes linguistiques ont une combinatoire beaucoup plus limitée que la plupart des autres signes de la langue (voir la section ?? sur Locutifs et interjections), mais ils n'en sont pas moins analysés de façon comparable et « Aïe! » n'a pas la même signification que « Aïaïaïe! » (prononcé a-ja-jaj)

4.4 Signifié, syntactique, signifiant

Les unités de première articulation, qui sont les portions de texte porteuses de signification, sont appelées les SIGNES LINGUISTIQUES. (Rappelons que nous désignons par *texte* aussi bien des textes écrits que des productions orales ou gestuelles.) La **signification** portée par un signe est appelée son SIGNIFIÉ. La **forme** d'un signe linguistique, en général un segment de texte, est appelée son SIGNIFIANT.

Les signes linguistiques se distinguent des signes de la plupart des autres systèmes sémiotiques par le fait qu'ils peuvent se combiner entre eux pour former de nouveaux signes qui expriment en général des sens qui ne peuvent être exprimés par aucun signe élémentaire. Les signes linguistiques ont donc un potentiel combinatoire qui fait qu'ils possèdent une troisième composante. Cette troi-

sième composante n'est pas réductible aux deux autres. Par exemple, les noms, lorsqu'ils se combinent à un article, déclenchent un accord en genre de cet article : *une* chaise, *un* fauteuil, *un* tabouret, *une* table, etc. Rien, ni dans la forme du nom, ni dans son sens, ne permet de prévoir quel sera cet accord — féminin ou masculin. Le locuteur ne peut utiliser correctement un nom en français que s'il a appris cette information en plus du signifiant et du signifié du nom.

Le signe linguistique possède donc une troisième composante, appelée son syntactique, qui contrôle sa **combinatoire** et qui n'est réductible ni aux propriétés du signifiant, ni à celles du signifié :

signe = < signifié, syntactique, signifiant >

Nous aurons l'occasion de décrire en détail le syntactique des signes. Celui-ci comprend plus d'informations qu'on ne le pense en général. Il comprend d'abord ce qu'on appelle la CATÉGORIE SYNTAXIQUE du signe. Pour un lexème, il s'agit de la PARTIE DU DISCOURS (verbe, nom, adjectif, etc.), mais aussi d'un certain nombre de traits qui contrôlent entre autres la combinatoire flexionnelle (comme le groupe de conjugaison pour les verbes) ou l'accord (comme le genre des noms). Le syntactique contient également la VALENCE, c'est-à-dire la liste des compléments régis par le signe avec le RÉGIME qui leur est imposé, c'est-à-dire les contraintes sur la nature du complément, sa place par rapport au gouverneur et les marques qui doivent l'accompagner (cas, préposition, etc.) (voir les Sections 3.3.10 sur Distribution et valence et 3.3.17 sur Tête interne et rection). Enfin, le syntactique comprend la description de la COOCCURRENCE LEXICALE RES-TREINTE, c'est-à-dire toutes les combinaisons qui sont lexicalement contraintes (par exemple le fait que amoureux s'intensifie par follement, blessé par grièvement, malade par gravement et improbable par hautement) (voir la section 6.11 sur Collocation et choix lié).

L'étude des signes linguistiques, en raison de leurs trois composantes, relève de trois disciplines différentes : l'étude des signifiés relève de la SÉMANTIQUE, l'étude du syntactique est le domaine de la SYNTAXE et l'étude des signifiants de la MORPHOLOGIE.

4.5 Le signe linguistique **Q**

La notion moderne de **signe linguistique** est généralement attribuée à Saussure (1916 : 98) :

check "Saussure" against "de Saussure" in whole book

« Le signe linguistique unit non une chose et un nom, mais un concept et une image acoustique. Cette dernière n'est pas le son matériel, chose purement physique, mais l'empreinte psychique de ce son, la représentation que nous en donne le témoignage de nos sens; elle est sensorielle, et s'il nous arrive de l'appeler « matérielle », c'est seulement dans ce sens et par opposition à l'autre terme de l'association, le concept, généralement plus abstrait. [...] Nous proposons de conserver le mot signe pour désigner le total, et de remplacer concept et image acoustique respectivement par signifié et signifiant; ces derniers termes ont l'avantage de marquer l'opposition qui les sépare soit entre eux, soit du total dont ils font partie. »

Saussure n'indique pas clairement s'il appelle signe linguistique n'importe quel segment de texte possédant un sens, comme nous le faisons dans cet ouvrage, ou si le terme s'applique uniquement aux unités minimales. Le paragraphe suivant sur l'arbitraire du signe concerne clairement les signes minimaux, appelés chez nous les morphèmes :

« Le lien unissant le signifiant au signifié est arbitraire, ou encore, puisque nous entendons par signe le total résultant de l'association d'un signifiant à un signifié, nous pouvons dire plus simplement : le signe linguistique est arbitraire. [...] Le mot arbitraire appelle aussi une remarque. Il ne doit pas donner l'idée que le signifiant dépend du libre choix du sujet parlant [...]; nous voulons dire qu'il est immotivé, c'est-à-dire arbitraire par rapport au signifié, avec lequel il n'a aucune attache naturelle dans la réalité. » (Saussure1916 : 101)

Chez Saussure, le signe linguistique est un élément à deux faces, signifié et signifiant. La troisième composante du signe, son **syntactique**, n'est pas explici-



(1)

tement considérée. Mel'čuk1993 est probablement un des premiers à insister sur la nécessité de cette composante dans la définition du signe linguistique.

4.6 Signifié ou exprimende

La terminologie usuellement utilisée pour désigner la forme et le contenu d'un signe linguistique — signifiant et signifié — véhicule une certaine conception du signe. Le signe est vu comme quelque chose qui signifie, c'est-à-dire quelque chose dont on saisit le signifiant pour accéder à son signifié. C'est une vision du rôle du signe que conteste vivement Lucien Tesnière1959 :

« Lorsque nous parlons, notre intention n'est pas de trouver après coup un sens à une suite de phonèmes qui lui préexistent, mais bien de donner une forme sensible aisément transmissible à une pensée qui lui préexiste et en est la seule raison d'être. En d'autres termes, le télégraphe est là pour transmettre les dépêches, non les dépêches pour faire fonctionner le télégraphe. »

Cela conduit Tesnière à inverser le point de vue et, pour mettre la terminologie en conformité avec la nécessité de partir du sens, à proposer de nommer EXPRIMENDE le contenu du signe et EXPRIMÉ la forme du signe. On souligne ainsi que le signe sert à exprimer une pensée et non pas à donner sens à une forme. Bien que tout à fait sensible aux arguments de Tesnière (voir la section 1.13 Parler et comprendre), nous continuerons à utiliser la terminologie de Saussure, qui est aujourd'hui universellement adoptée.

4.7 Sémantème, syntaxème, morphème

Le fait que le signe ait trois composantes implique qu'il y a trois façons de l'appréhender et autant de façons de définir les signes minimaux : les signes minimaux du point de vue du sens seront appelés les sémantèmes, les signes minimaux du point de vue de la combinatoire les syntaxèmes et les signes minimaux du point de vue de la **forme** les MORPHÈMES.

Le grand problème de la caractérisation des signes minimaux vient de la noncorrespondance entre les unités minimales de forme et les unités minimales de sens. Par exemple, dans Aya dort à poings fermés, le segment à poings fermés est une unité minimale de sens exprimant l'intensification du verbe dormir. Ce segment correspond à un choix unique et indivisible du locuteur et pourtant cette unité est construite en utilisant d'autres unités dont on peut entrapercevoir la contribution, puisqu'il s'agit d'un emploi métaphorique figé de la combinaison

libre à poings fermés 'en ayant les poings fermés'. Un autre exemple : le mot décapsuleur désigne un type particulier d'objet et est donc un choix unique et indivisible fait pour désigner un tel objet. Ce mot n'en est pas moins l'assemblage de trois unités, $d\acute{e}+capsul+eur$, que l'on retrouve dans d'autres mots (capsule, $d\acute{e}capsuler$, $d\acute{e}fibrillateur$, etc.). Les signes A poings fermés ou decapsuleur sont des sémantèmes. Les composantes $d\acute{e}$ -, capsule et -eur de $d\acute{e}capsuleur$ sont des morphèmes : il s'agit bien de signes, car on peut leur attribuer une signification et déterminer leur contribution au sens de $d\acute{e}capsuleur$, et il s'agit de signes minimaux, car on ne peut les décomposer à nouveau en des formes qui possèdent une signification.

Cette non-correspondance entre unités minimales de sens et de forme vaut aussi pour les unités minimales de combinatoire. Si un sémantème comme décapsuleur se comporte comme un nom simple — par exemple bol ou couteau —, un sémantème comme Àvoir les pieds sur terre (au sens de 'avoir le sens des réalités') ne se comporte pas comme un verbe simple, mais plutôt comme une combinaison libre du verbe avoir avec des compléments — par exemple avoir les mains sur la table. Le verbe avoir se conjugue de la même façon dans les deux cas, il se combine de la même façon avec la négation, etc. Ceci a notamment pour conséquence que le sémantème est séparable : Il n'a vraiment pas les pieds sur terre. Ainsi le composant avoir du sémantème Àvoir les pieds sur terre. Ainsi le vue de sa combinatoire le même verbe avoir que celui qui s'utilise librement dans avoir les mains sur la table. C'est une telle unité que nous appelons un syntaxème. Les autres composantes du sémantème Àvoir les pieds sur terre sont aussi des syntaxèmes : le, pied, sur, terre, ainsi que les syntaxèmes flexionnels qui marquent le nombre.

Par contre, le morphème *capsule* de DÉCAPSULEUR n'est pas un syntaxème : *capsule* dans DÉCAPSULEUR n'a plus la combinatoire d'un nom. C'est *décapsuleur* comme un tout qui rentre dans des combinaisons et, par exemple, qui impose les accord en genre (*un décapsuleur blanc* vs *une capsule blanche*).

4.8 L'identification des unités de la langue

Dans toute description analytique d'une chose, qu'il s'agisse d'un être vivant ou d'une machine, on cherche à identifier les parties constituantes. Pour qu'une partie de la chose soit reconnue comme un constituant, on doit pouvoir identifier sa contribution au tout. On doit pouvoir en délimiter les contours (le signifiant), en déterminer les possibilités de combinaison avec les autres éléments du système (le syntactique) et en identifier la contribution au fonctionnement



général du système (le signifié). Il n'est réellement intéressant d'isoler un constituant que si l'on peut l'extraire pour le réutiliser ailleurs. Ceci est particulièrement vrai pour les signes de la langue. Ceux-ci ne deviennent identifiables que s'ils sont utilisés plusieurs fois dans des contextes différents et que dans ces mêmes contextes d'autres signes sont utilisées. On peut alors les extraire pour les recombiner autrement et c'est ce qui fait l'infinie richesse de la langue.

Classifier les différents morceaux en fonction des éléments qui peuvent les remplacer (les RAPPORTS PARADIGMATIQUES) et des éléments avec lesquels ils peuvent se combiner (les rapports syntagmatiques) s'appelle l'Analyse dis-TRIBUTIONNELLE. La technique qui consiste à remplacer un segment par un autre pour voir s'il s'agit d'un signe s'appelle le PRINCIPE DE COMMUTATION. Un ensemble de segments qui peuvent commuter les uns avec les autres s'appelle un PARADIGME DE COMMUTATION.

L'analyse distributionnelle s'est d'abord intéressée aux unités minimales de forme, en délaissant volontairement le sens (voir l'encadré qui suit) et sans clairement distinguer le syntactique du signifiant. Nous appliquerons pour notre part l'analyse distributionnelle aux trois composantes du signe linguistique. Nous considérons en particulier que le sens est un observable dans la mesure où il est possible d'une part d'observer dans quel contexte d'énonciation un énoncé est produit et donc de faire des hypothèses sur les intentions du locuteurs qui le produit. Et d'autre part, de demander à un locuteur quels sont les énoncés qu'il aurait été possible de produire pour obtenir le même résultat, c'est-à-dire les énoncés qui possèdent le même sens (voir l'encadré ?? sur Les observables : textes et sens).

Les deux chapitres suivants seront consacrés aux deux types extrêmes de signes minimaux : les morphèmes et les sémantèmes. Les syntaxèmes seront introduits aussi dans cette partie, mais leur étude approfondie aura lieu dans la Partie 3.

4.9 Structuralisme et distributionnalisme



L'identification des phonèmes et des signes linguistiques d'une langue ne va pas de soi. Le courant STRUCTURALISTE, qui s'est développé à la suite de la publication du Cours de linguistique générale de Ferdinand de Saussure, propose de prendre comme seule base d'analyse l' « observable » de la langue, c'est-àdire les énoncés produits par les locuteurs ou acceptables pour eux, et élabore un certain nombre de méthodes basées sur la manipulation de ces énoncés. Suivant la tradition de la PSYCHOLOGIE BEHAVIORISTE, le sens des énoncés, considéré comme non observable par essence, est utilisé aussi peu que possible : la langue, comme d'autres comportements (angl. *behaviors*) humains, est analysée de l'extérieur, sans faire d'hypothèse a priori sur ce qui peut se passer dans le cerveau du locuteur.

Le stade le plus abouti de cette approche basée uniquement sur l'observation des énoncés a été atteint pas l'école distributionnaliste, dont les principaux acteurs ont été Leonard Bloomfield, Edward Sapir, Benjamin L. Whorf, Charles F. Hockett et Zelig Harris. Ce courant est né de l'étude des langues amérindiennes, langues essentiellement orales et en voie de disparition pour la plupart, pour lesquelles il n'y avait ni écriture, ni description préalable d'aucune sorte. La simple transcription des énoncés produits par les locuteurs nécessitait d'établir très rapidement une liste de phonèmes et une segmentation en morphèmes et en mots quelque peu raisonnable.

L'influence du distributionnalisme s'est estompée dans les années 1960 avec l'émergence de la « nouvelle syntaxe » de Noam Chomsky. En pointant les limites de l'analyse de corpus pour la compréhension du fonctionnement de la langue, il rétablit l'introspection comme méthodologie acceptable pour la linguistique, faisant ainsi perdre à la syntaxe sa base empirique. Seuls les besoins du traitement automatique de la langue et la numérisation des textes pour le web depuis les années 1990 ont poussé les syntacticiens à se réintéresser à l'adaptation de leurs analyses aux données observables.



Exercices

Exercice 1 Lorsqu'un professeur veut faire taire un élève, il a plusieurs possibilités à sa disposition : (1) il peut dire « *Taisez-vous*! » (2) ou « *Chut*! », (3) ou produire un son particulier « *Tttt* » avec les lèvres en avant et un claquement de langue à l'avant du palais, (4) ou encore lever l'index ou (5) ou jeter un regard accusateur. De ces cinq stratégies, lesquelles utilisent des signes ? Lesquels sont des signes linguistiques ? Lesquels ne vérifient pas la double articulation ?



Qu'est-ce qui distingue un idéogramme d'une lettre de l'al-
Quelles sont les trois composantes du signe linguistique?
Pourquoi distinguons-nous trois types d'unités minimales?
Quel est le principe de base de l'analyse distributionnelle?



Lectures additionnelles

Le livre de lexicologie d'Alain **Polguère2003** est un bon complément à notre ouvrage de syntaxe. On y trouvera une introduction à la fois simple et riche du signe linguistique et une présentation de la structure du lexique et du sens lexical. L'encyclopédie des sciences du langage de **DucrotSchaeffer1995** est un ouvrage que nous recommandons particulièrement, notamment le chapitre consacré au « Signe » (p. 253 à 256), qui situe le système linguistique parmi les systèmes sémiotiques en général.

Nous avons déjà parlé des ouvrages de Saussure1916, Bloomfield1933 et Humboldt1836 au Chapitre 1. Concernant les distributionnalistes, on pourra consulter Sapir1921 et Harris1951. Nous recommandons particulièrement la lecture de Hockett1958. Nous avons déjà parlé d'Edward Sapir au Chapitre 2 et nous reparlerons de Charles F. Hockett dans les chapitres 3.1 et 3.4.

Oswald Ducrot et Jean-Marie Schaeffer1995 Nouveau dictionnaire encyclopédique des sciences du langage, Editions du Seuil.

Zellig S. Harris1951 Methods in structural linguistics, University of Chi-



cago Press.

Alain Polguère2003 Lexicologie et sémantique lexicales. Notions fondamentales. Presses de l'Université de Montréal.

Edward Sapir1921 Language: An introduction to the study of speech, New York.



Citations originales

Citations de la section 4.2.

Humboldt (Humboldt1836: 60):

Die Absicht und die Fähigkeit zur Bedeutsamkeit [...] macht allein den artikulierten Laut aus, und es lässt sich nichts andres angeben, um seinen Unterschied auf der einen Seite vom tierischen Geschrei, auf der anderen vom musikalischen Ton zu bezeichnen.



Corrections des exercices

Corrigé 1 Les trois premiers sont des signes linguistiques et les deux premiers, « *Taisez-vous*! » et « *Chut*! », vérifient la double articulation et se décomposent en phonèmes. Néanmoins *Chut* n'a pas de combinatoire syntaxique (voir la section ?? sur *Locutifs et interjections*). Le signe linguistique *Tttt* (dont on ne sait pas très bien comme l'orthographier) appartient à un sous-système hors du système phonologique du français, bien que ce soit un signe du français (qui n'existe pas a priori dans d'autres langues). Le doigt levé est aussi un signe conventionnel, comme le sont les applaudissements. Le regard accusateur peut aussi être considéré comme un signe et il est probablement plus conventionnel qu'on ne le pense a priori : pas sûr que le même regard soit interprété de la même façon dans une autre communauté.

Corrigé 2 Un idéogramme correspond à un morphème, alors qu'une lettre de l'alphabet est la transcription d'un phonème.

Corrigé 3 Voir la section 4.4 Signifié, syntactique, signifiant.

Corrigé 4 À chacune des trois composantes du signe correspond un découpage particulier. Voir la section 4.7 Sémantème, syntaxème, morphème.

Corrigé 5 Le principe de base de l'analyse distributionnelle est le principe de commutation, comme nous allons le voir au chapitre suivant.

5 Morphèmes et syntaxèmes : Unités de forme vs unités de combinatoire

5.1 Principe de commutation

Le principe de commutation est le principe de base de l'ANALYSE DISTRIBU-TIONNELLE (voir section 4.8 sur *L'identification des unités de la langue*). Il consiste à utiliser la commutation comme mode de décomposition des signes linguistiques.

Une commutation est le remplacement d'un segment de texte par un autre. Par exemple, on peut passer de *construction* à *destruction* en commutant *con-*/kɔ-/ avec *de-*/de-/. Une commutation a toujours lieu dans un ENVIRONNEMENT donné : ici l'environnement considéré est ___ struction.

La commutation d'un élément permet de dissocier cet élément de son environnement. La commutation précédente permet de postuler la décomposition suivante : *construction* = *con* + *struction*.

Le symbole + désigne une opération purement formelle de décomposition des textes que l'on appelle la concaténation et qui est ainsi l'opération de combinaison des signifiants la plus élémentaire.

Nous nous limitons pour l'instant à la combinaison des signifiants par concaténation (les signifiants sont collés les uns à la suite des autres), mais on verra par la suite qu'il existe des signes qui ont des signifiants discontinus et d'autres dont les signifiants fonctionnent comme des opérateurs (voir l'alternance dans la section 5.22 sur *Amalgame*, *alternance et mégamorphe* et la troncation dans l'encadré ?? sur *Syntaxème zéro et troncation en français*).

5.2 Commutation et exclusion mutuelle

Pour considérer qu'il y a vraiment commutation, il faut remplir une condition supplémentaire qu'on appelle l'exclusion mutuelle. Considérons des énoncés comme *Pierre vient demain* et *Pierre vient en voiture*. On ne veut pas dire ici que *demain* et *en voiture* commutent dans l'environnement « *Pierre vient* —__-», car les deux compléments sont cumulables (*Pierre vient en voiture demain*) et qu'aucun n'est obligatoire (*Pierre vient*). Il est donc clair que *demain* et *en voiture* n'occupent pas à proprement parler la même place et ne se remplacent pas l'un l'autre.

Nous dirons donc que deux signes A et B COMMUTENT dans l'environnement X_Y s'ils **peuvent se remplacer** l'un l'autre et qu'en plus ils **s'excluent** l'un l'autre, c'est-à-dire que XAY et XBY sont acceptables et que XABY et XBAY sont inacceptables.

Par exemple, A = pas et B = jamais commutent dans l'environnement *Pierre ne vient* <u>en voiture</u>. Ou encore A = -ait et B = -era commutent dans l'environnement *chant*

L'exclusion mutuelle joue un grand rôle théorique, car elle permet de postuler des positions structurales : deux éléments qui s'excluent mutuellement occupent généralement la même position structurale, puisqu'ils se « remplacent » l'un l'autre. Lorsque la structure s'entend en termes de combinaisons libres, nous parlerons de structure syntaxique et positions syntaxiques. Lorsque la structure s'entend en termes de précédence et de contiguïté, c'est-à-dire d'ordre linéaire, nous parlerons de structure topologique et positions topologiques.

Un ensemble de signes qui peuvent commuter les uns avec les autres est appelé un paradigme de commutation.

Une position structurale est caractérisée par le paradigme des éléments qui peuvent occuper cette position.

Un couple d'exemples qui se caractérise par une commutation minimale est appelé une PAIRE MINIMALE.

Par exemple, u = *Pierre ne vient pas en voiture* et v = *Pierre ne vient jamais en voiture* constituent une paire minimale : on passe de u à v par la commutation de *pas* et *jamais* et cette commutation est minimale, car on ne pourrait faire une commutation raisonnable sur un segment plus petit que *pas* ou *jamais*. Le couple *chantait* vs *chantera* n'est pas une paire minimale, car il possible de faire une commutation sur un segment plus petit et de mettre en rapport *chantait* avec *chanterait* et *chanterait* avec *chantera.*

5.3 Décomposition propre des signes

Le fait qu'une commutation soit possible n'assure pas que les unités dégagées soient des signes linguistiques; il faut pour cela des conditions supplémentaires. Si la commutation entre *con-* et *de-* dans la paire minimale *construction* vs *destruction* est linguistiquement intéressante, c'est parce que *construction* et *destruction* sont deux signes sémantiquement liés et qu'on peut postuler que la commutation de *con-* et *de-* au niveau des signifiants est corrélée au changement de sens au niveau des signifiés. L'existence d'autres paires comme *constitution* et *destitution* et donc d'une décomposition potentielle *con + stitution* permet de renforcer le statut de *con-* /kɔ-/ comme unité de la langue. Mais pour que *con-* soit réellement considéré le signifiant d'un signe, il aurait fallu que sa contribution sémantique soit équivalente dans *construction* et *constitution*, ce qui n'est pas la cas.

Nous allons ajouter une condition qui assure qu'un segment mis à jour par une

commutation possède un signifié et définir un **principe de commutation enrichi** qui prend en compte le signe dans sa globalité, signifié et syntactique compris.

Le signe X se décompose proprement en A + B si X = A + B, s'il existe des segments A' et B' tels A + B', A' + B et A' + B' soient des signes et si A + B est à A' + B ce que A + B' est à A' + B' (ou, ce qui revient plus ou moins au même, si A + B est à A' + B').

Nous notons \boxplus l'opération de combinaison propre. La notation « $X = A \boxplus B$ » signifie donc que X se décompose proprement en A et B.

Quand nous exigeons que A + B soit à A' + B ce que A + B' est à A' + B', nous mesurons aussi bien la **différence de signifiants** entre A + B et A' + B que la **différence de syntactiques ou de signifiés**.

Prenons quelques exemples. Le signe X = un chat se décompose proprement en $A \boxplus B$ avec A = un et B = chat. On peut en effet faire commuter A avec A' = le et B avec B' = chien; les quatre combinaisons sont des signes (un chat, le chien, un chien) et un chat est à le chat ce que un chien est à le chien. Ou encore un chat est à un chien ce que le chat est à le chien.

Le signe X = avançait se décompose proprement en $A \boxplus B$ avec A = avanç- et B = -ait. On peut en effet faire commuter A avec A' = recul- et B avec B' = -era; les quatre combinaisons sont des signes (avançait, avancera, reculait, reculera) et avançait est à avancera ce que reculait est à avancera.

Le signe X = broyeur se décompose proprement en $A \boxplus B$ avec A = broy- et B = -eur. On peut en effet faire commuter A avec A' = compress- et B avec B' = -ons; les quatre combinaisons sont des signes (*broyeur*, *broyons*, *compresseur*, *compressons*) et *broyeur* est à *broyons* ce que *compresseur* est à *compressons*.

Un cas particulier important est celui où A existe en tant que signe autonome et l'on peut donc prendre pour B' un segment vide. On a alors :

Le signe X se décompose proprement en $A \boxplus B$ si X = A + B, si A est un signe et s'il existe un autre signe A' tel que A' + B soit un signe et que A + B soit à A' + B ce que A est à A' (ou, ce qui revient plus ou moins au même, que A + B soit à A ce que A' + B est à A').

Ainsi le signe X = facilement se décompose proprement en $A \boxplus B$ avec A = facile et B = -ment, puisqu'on peut faire commuter A avec A' = utile et que facilement est à facile ce que utilement est à utile.

Figure 5.1

On peut représenter l'application du principe de commutation enrichi par un rectangle analogique. Les quatre combinaisons A+B, A+B', A'+B et A'+B' occupent les quatre coins du rectangle et les côtés opposés du rectangle indiquent des rapports de proportionnalité équivalents :

Un signe linguistique qui se décompose proprement est dit DIAGRAMMA-TIQUE.

On peut voir la DIAGRAMMATICITÉ comme la possibilité d'être décomposé par un diagramme tel qu'un rectangle analogique. Un signe diagrammatique est nécessairement construit à partir d'autres signes et sa constructionalité est visible (ICONIQUE dirait le logicien américain Charles S. Peirce, qui est à l'origine d'une grande partie de la terminologie utilisée aujourd'hui en sémiotique et notamment du terme *diagramme*).

Lorsqu'un signe X se décompose proprement en $A \boxplus B$, on peut attribuer des sens à A et B de telle façon que le sens de X se calcule de manière **relativement compositionnelle** à partir des sens de A et de B. On en déduit que :

Si le signe X se décompose proprement en $A \boxplus B$, alors A et B peuvent être considérés comme des SIGNES.

En effet, comme la commutation est propre, on peut considérer que la contribution de A est la même dans A + B et A + B' et évaluer cette contribution, notamment lorsqu'on connaît déjà le sens de B et B'.

Les composants d'une décomposition propre ne préexistent pas nécessairement à la décomposition. C'est parce qu'une décomposition potentielle vérifie le principe de commutation enrichi que cette décomposition est propre et que les composants peuvent être reconnus comme des signes. Prenons l'exemple de broyeur. Pour l'apprenant du français, le signe -eur ne préexiste pas à la décomposition des mots de type broyeur et compresseur. C'est parce qu'un broyeur est à broyons ce qu'un compresseur est à compressons que l'on peut décomposer broyeur et donner un sens à -eur (un X-eur est une machine qui sert à X-er) et considérer que broyeur = broy \boxplus eur (voir discussion dans l'encadré ?? sur La quatrième proportionnelle).

La conséquence de la remarque précédente est que les signes ne sont pas des éléments donnés a priori, des briques qui préexisteraient à la construction des énoncés. C'est au contraire des énoncés auxquels nous sommes confrontés que nous extrayons ces briques, que nous pouvons ensuite assembler autrement pour construire de nouveaux énoncés.

Attention : on ne confondra pas la diagrammaticité avec la compositionalité. Nous donnerons au terme compositionalité une acception beaucoup plus restrictive (voir la section 6.3 Chaque sémantème suppose un choix). La diagrammaticité est une notion un peu floue et graduelle, comme l'est la notion de décomposition propre. Le sens que nous attribuons à un signe comme -eur est vague. Un compresseur n'est pas n'importe quel appareil qui sert à compresser (un compresseur compresse de l'air et pas autre chose) et c'est en ce sens que la combinaison compress \boxplus eur n'est pas stricto sensu compositionnelle. La diagrammaticité est une forme faible de la compositionalité.

5.4 La quatrième proportionnelle 🍳

L'énonciation du **principe de commutation enrichi** est postérieure aux travaux de Saussure. Mais elle est sous-jacente à la notion fondamentale de RAP-



PORTS ASSOCIATIFS, que Saussure (Saussure1916 : 177, 179) oppose aux rapports SYNTAGMATIQUES, et que l'on nomme aujourd'hui, à la suite de Louis Hjelmslev et de Roman Jakobson, les rapports paradigmatiques :

« En dehors du discours, les mots offrant quelque chose de commun s'associent dans la mémoire, et il se forme ainsi des groupes au sein desquels règnent des rapports très divers. Ainsi le mot enseignement fera surgir inconsciemment devant l'esprit une foule d'autres mots (enseigner, renseigner, etc., ou bien armement, changement, etc., ou bien éducation, apprentissage); par un côté ou un autre, tous ont quelque chose de commun entre eux. [...] Le rapport syntagmatique est *in praesentia*; il repose sur deux ou plusieurs termes également présents dans une série effective. Au contraire le rapport associatif unit des termes in absentia dans une série mnémonique. [...]

Ouand quelqu'un dit *marchons!*, il pense inconsciemment à divers groupes d'associations à l'intersection desquels se trouve le syntagme marchons! Celui-ci figure d'une part dans la série marche! marchez!, et c'est l'opposition de marchons! avec ces formes qui détermine le choix; d'autre part, marchons! évoque la série montons! mangeons! etc., au sein de laquelle il est choisi par le même procédé; dans chaque série, on sait ce qu'il faut faire varier pour obtenir la différenciation propre à l'unité cherchée. »

Bien qu'il n'ait pas explicitement introduit le RECTANGLE ANALOGIQUE (que certains attribuent au typologue Joseph Greenberg), Saussure (Saussure1916: 228, 231) montre le rôle de la OUATRIÈME PROPORTIONNELLE dans la construction de nouvelles unités significatives :

« Toute création analogique peut être représentée comme une opération analogue au calcul de la quatrième proportionnelle. [...] Magasinier n'a pas été engendré par magasin; il a été formé sur le modèle de prisonnier: prison, etc. [...] Pour former indécorable, nul besoin d'en extraire les éléments (indécor-able); il suffit de prendre l'ensemble et de le placer dans l'équation : pardonner: impardonnable, etc., = décorer: x.

```
x = indécorable. [...]
```

Sur le modèle de pension : pensionnaire, réaction : réactionnaire, etc., quelqu'un peut créer interventionnaire ou répressionnaire, signifiant « qui est pour l'intervention », « pour la répression » :

```
réaction : réactionnaire = répression : x.
```

```
x = répressionnaire. [...]
```

À tout instant on rencontre des combinaisons sans lendemain que la langue n'adoptera probablement pas. Le langage des enfants en regorge, parce qu'ils connaissent mal l'usage et n'y sont pas encore asservis; ils disent *viendre* pour *venir*, *mouru* pour *mort*, etc. Toutes ces innovations sont en soi parfaitement régulières; elles s'expliquent de la même façon que celles que la langue a acceptées; ainsi *viendre* repose sur la proportion :

 $\acute{e}teindrai: \acute{e}teindre = viendrai: x.$

x = viendre. »

Il est important de noter qu'il y a chez Saussure une très grande réticence à considérer les éléments qui commutent comme des signes : « Pour former *indécorable*, nul besoin d'en extraire les éléments (*in-décor-able*) ». Comme il le montre, il n'est effectivement pas nécessaire d'extraire les composants pour expliquer la formation de nouvelles unités significatives. Aujourd'hui encore une telle décomposition (et la notion de *morphème* qui en découle) est mise en question. La décomposition des signes en signes non autonomisables (c'est-à-dire qui ne peuvent être prononcés seuls et être interprétés seuls) est d'une certaine façon une abstraction, bien utile cependant pour la modélisation des langues. Ce ne sont pas des observables au sens strict, mais des constructions théoriques.

5.5 Quasi-signes

Si nous reprenons notre exemple initial, on peut placer les mots construction, destruction, constitution et destitution aux quatre coins d'un rectangle, mais il n'y a pas de rapports de proportionnalité entre les côtés de ce rectangle : construction n'est pas à destruction ce que constitution est à destitution. En fait, il ne semble pas qu'il existe de paire qui permette de former un rectangle analogique avec la paire construction-destruction. Plus généralement, il n'existe pas d'éléments A' et B' qui permettent de valider une décomposition propre de con + struction. La combinaison con + struction est donc peu diagrammatique. La diagrammaticité est une notion graduelle et on peut dire que la combinaison con + stitution est encore moins diagrammatrique, car il y a peu de rapport de sens entre constitution et les autres mots en -stitution (destitution, institution, restitution, prostitution). Pour les formes qui interviennent dans une décomposition impropre, comme struction ou stitution, nous introduisons le terme quasi-signe.

Un QUASI-SIGNE est un segment qui apparaît dans un paradigme de commutation, mais pour lequel la commutation n'est pas (suffisamment) propre.

Le quasi-signe possède un signifiant clair, délimité par le principe de commutation, mais ne possède pas de signifié à proprement parler.

Nous notons + l'opération de combinaison entre les quasi-signes et nous nommons cette opération, par opposition à l'opération ⊞, l'opération DE COMBINAISON IMPROPRE.

La décomposition des signes lorsqu'elle est peu diagrammatique est probablement peu visible au locuteur ordinaire, car elle ne joue pas de rôle en synchronie. Elle permet seulement d'avoir des intuitions sur la parenté et l'origine des mots. Par exemple, il est peu visible que arriver et dériver sont des mots construits avec les éléments a-, dé- et rive, sur le même schéma que aborder ou atterrir ou que déplacer ou déterrer. Le principe de commutation enrichi s'applique mal à nouveau : si on considère A = a-, $A' = d\acute{e}$ -, B = riv(er) et B' = port(er), et qu'on les combine, on obtient bien quatre signes arriver, dériver, apporter, déporter, mais arriver n'est pas exactement à dériver ce que apporter est à déporter. Il y a néanmoins entre arriver et apporter ou entre dériver et déporter ou encore entre arriver et dériver des liens sémantiques. Ces liens permettent d'entrevoir derrière les sens réels de arriver et dériver des lectures « compositionnelles » de a- + riv(er) 's'approcher de la rive' et de dé- + riv(er) 's'éloigner de la rive'. Les affixes a- et dé- s'utilisent pour qualifier des mouvements dans un sens ou un autre (de même que dans les prépositions à et de qui sont aussi les marqueurs de tels mouvements : il part à la plage, il part **de** la plage). En conclusion, la décomposition de arriv(er) en a-+riv(er) ne donne pas une combinaison propre, mais elle garde une certaine réalité en synchronie par la proximité des formes a- et riv(er) avec d'autres formes qui interviennent dans des signes sémantiquement liés.

5.6 Combinaison libre

Nous avons défini le principe de commutation enrichi à partir d'une commutation sur chacune des deux parties de la combinaison A + B (commutation de A' avec A, commutation de B' avec B). Bien entendu, plus le paradigme de commutation est important, plus la combinaison est considérée comme régulière. Nous allons maintenant voir à quelle condition une combinaison relève de la syntaxe.

Si l'on considère le mot *blessure*, on voit que *bless*- commute proprement avec *brûl*-, *griff*-, *piqu*-, *égratign*-, *cass*-, *fêl*-, ... et que par ailleurs -*ure* commute proprement avec -*ons*, -*ez*, *ais*, -*a*, ... et que toutes les combinaisons deux à deux sont valides (*brûlez*, *griffais*, ...).

Venons-en au point qui nous intéresse : la combinaison $bless \boxplus ure$ est pourtant bien différente de la combinaison $bless \boxplus ons$. Comment caractériser cette différence?

Les signes qui commutent proprement avec *bless*- dans *blessons* sont les mêmes que ceux qui commutent avec *bless*- dans *blessez* ou *blessais*. Il s'agit d'une classe que l'on va d'ailleurs retrouver dans de nombreux environnements : la classe des verbes. Rien de cela avec *blessure* : les éléments qui commutent avec *bless*- dans *blessure* forment une classe atypique qui contient un certain nombre de verbes, mais aussi des quasi-signes comme *struct*- ou *fract*-. Autrement dit, aucun autre signe n'a la même distribution que *-ure*

Nous appelons distribution d'un signe l'ensemble des environnements dans lequel ce signe peut apparaître.

Plus généralement, un **ensemble d'environnements compatibles** avec un ensemble donné de **signes** est appelée une famille d'environnements. La distribution d'un signe est donc la famille d'environnements de ce signe.

La notion de distribution n'est pas univoque : elle dépend de la taille des environnements considérés et de la nature des environnements considérés. Nous nous intéressons ici à des environnements étroits, de l'ordre du mot. Nous élargirons l'empan de nos environnements dans le Chapitre ?? sur les *Catégories microsyntaxiques*. Par ailleurs, nous ne considérons pour l'instant que la dimension linéaire du texte. Nous verrons dans la partie 3 comment définir une structure syntaxique et c'est sur les environnements structurels que nous appliquerons notre analyse distributionnelle du Chapitre ??. Autrement dit, nos environnements seront des portions de la structure et non des segments de texte.

Revenons à la distribution de *-ure*. Cette distribution n'est pas déductible simplement des distributions d'autres signes, par exemple par l'intersection ou la réunion de familles d'environnements d'autres signes (voir l'encadré ?? sur la *Théorie des ensembles*; nous pouvons nous contenter pour l'instant d'une compréhension superficielle de ces notions.). Nous dirons que la distribution de *-ure* est IRRÉGULIÈRE.

On peut inverser le point de vue et voir -*ure* comme un environnement.

Un ensemble de signes compatibles avec un ensemble donné d'environnements est appelé une classe distributionnelle.

Dire que la distribution de *-ure* est irrégulière revient à dire que la classe des signes qui peuvent se combiner avec *-ure* est irrégulière.

On dit que A commute librement dans la combinaison propre $A \boxplus B$ si la classe des signes qui **commutent proprement** avec A est assez **régulière** et peut notamment se déduire des classes de signes qui commutent avec A dans d'autres combinaisons.

Dans l'énoncé *Le chat avançait*, les quatre sous-segments *le, chat, avanç-, -ait* sont des unités syntaxiques. Chaque segment est bien le signifiant d'un signe comme le montre les commutations propres suivantes :

Table 5.1

le	chat	avanç-	-ait	
un	chien	grogn-	-e	
mon	garçon	chant-	-era	
ce	public	march-	-a	

Ces commutations sont libres, car chaque élément que nous avons introduit possède à peu près la même distribution que ceux avec lesquels ils commutent. En particulier, toutes les combinaisons qui résultent des commutations considérées sont valides : *le public grogne, ce garçon avancera, mon chat chanta*, etc.

On dit que A et B se combinent librement dans la combinaison propre $A \boxplus B$.



La notation « $X = A \oplus B$ » signifie que X est le résultat de la combinaison libre de A et B. On appelle cette opération ⊕ l'opération de combinaison LIBRE.

La combinaison *bless* \oplus *ons* est libre, tandis que la combinaison *bless* \boxplus *ure* est propre, mais LIÉE.

Un signe qui commute librement avec son environnement est appelé une UNITÉ SYNTAXIQUE. (Voir Chapitre 7 pour une définition plus précise.)

5.7 Liberté de combinaison et opposition parole/langue



L'opposition entre combinaison libre et liée marque la frontière entre ce que Saussure nomme la langue et la parole. Les combinaisons libres font partie de la PAROLE, c'est-à-dire des productions que chaque locuteur est libre de créer à sa guise. Les combinaisons liées, au contraire, font partie de la LANGUE, c'està-dire des connaissances partagées par les locuteurs d'une même langue et qui constituent leur stock lexical. Saussure (Saussure1916: 172) ne dit pas autre chose dans l'extrait suivant (même si la notion de combinaison libre et liée n'est pas définie formellement):

« Le propre de la parole, c'est la liberté des combinaisons. [C'est nous qui soulignons]

On rencontre d'abord un grand nombre d'expressions qui appartiennent à la langue; ce sont les locutions toutes faites, auxquelles l'usage interdit de rien changer, même si on peut distinguer, à la réflexion, des parties significatives. [...]

Mais ce n'est pas tout; il faut attribuer à la langue, non à la parole, tous les types de syntagmes construits sur des formes régulières. [...] Quand un mot comme indécorable surgit dans la parole, il suppose un type déterminé, et celui-ci à son tour n'est possible que par le souvenir d'un nombre suffisant de mots semblables appartenant à la langue (impardonnable, intolérable, infatigable, etc.). Il en est exactement de même des phrases et des groupes de mots établis sur des patrons réguliers; des combinaisons comme la terre tourne, que vous dit-il? etc., répondent à des types généraux, qui ont à leur tour leur support dans la langue sous forme de souvenirs concrets.

Mais il faut reconnaître que dans le domaine du syntagme, il n'y a pas de limite tranchée entre le fait de langue, marque de l'usage collectif, et le fait de parole, qui dépend de la liberté individuelle. Dans une foule de cas, il est difficile de classer une combinaison d'unités, parce que l'un et l'autre facteurs ont concouru à la produire, et dans des proportions qu'il est impossible de déterminer »

5.8 Théorie des ensembles

Pour mieux comprendre la suite, il ne sera pas inutile de rappeler les bases de la théorie des ensembles. Un ENSEMBLE est composé d'objets qui sont appelés ses ÉLÉMENTS. Par exemple, l'alphabet latin moderne est un ensemble comportant 26 éléments, que l'on appelle des lettres.

La relation qui lie un élément avec l'ensemble qui le contient est appelée la relation d'APPARTENANCE et est notée €. Par exemple, le lexème MANGER appartient à l'ensemble V des verbes du français :

(1) $manger \in V$.

On peut définir un ensemble de deux façons :

- soit en extension, en donnant la liste de ses éléments entre accolades : V = { manger, chanter, dormir, lire, ... };
- soit en Intension, par une définition formelle comme une formule ou une opération : par exemple, V est l'ensemble des syntaxèmes qui peuvent se combiner avec les syntaxèmes flexionnels de temps.

Il existe un unique ENSEMBLE VIDE, noté \emptyset , qui est donc le seul ensemble sans éléments.

On définit une relation entre les ensembles appelée l'inclusion et notée \subseteq . L'ensemble A est dit **inclus** dans l'ensemble B (et on note $A \subseteq B$) si tous les éléments de A sont dans B. On dit encore que A est un sous-ensemble de B ou une partie de B.

 $\begin{array}{cc} \text{(4)} & A \subseteq B \ \complement_A B \\ & \text{Inclusion Complémentaire} \end{array}$

L'inclusion est une relation d'ordre (voir la section 9.31 Dépendance, dominance et transitivité pour une définition formelle) sur les ensembles, similaire à la relation d'ordre \leq sur les nombres (d'où la notation \subseteq). Dire « $A \subseteq B$ » revient à dire que A est plus petit que B pour la relation d'inclusion. Néanmoins contrairement à la relation \leq , l'inclusion est un ordre partiel, puisque deux ensembles peuvent ne pas être ordonnés l'un par rapport à l'autre, notamment lorsqu'ils sont disjoints (c'est-à-dire n'ont pas d'éléments en commun).

À tout ensemble B inclus dans A, on peut associer le complémentaire ${\tt C}_A{\tt B}$ de B dans A.

À tout ensemble E, on peut associer l'ensemble des parties de E, noté 2^{E} ou PPPFIXME(E), avec un PPPFIXME comme *partie*. La notation 2^{E} repose sur le fait que si E a n éléments, alors 2^{E} a 2^{n} éléments.

Les ensembles \emptyset et E sont respectivement le plus petit et le plus grand élément de 2^E pour l'inclusion.

On ne confondra pas l'élément x de E avec le SINGLETON $\{x\}$, qui est un élément de 2^{E} .

On peut définir sur 2^E deux opérations binaires, l'union et l'intersection. L'union de A et B, notée $A \cup B$, est l'ensemble qui contient à la fois les éléments de A et ceux de B, tandis que l'intersection de A et B, notée $A \cap B$, est l'ensemble des éléments communs à A et B. On peut visualiser ces ensembles sur les schémas suivants :

(5) $A \cup B A \cap B$ Union Intersection

L'ensemble 2^E muni des opérations binaires \cap et \cup et de la relation d'ordre \subseteq possède d'excellentes propriétés similaires à celles de l'ensemble des nombres naturels muni des opérations + et \times et de la relation \le . En un sens, ces opérations et relations « structurent » l'ensemble 2^E ; une telle structure est appelée une STRUCTURE ALGÉBRIQUE.

5.9 Dualité 🕏



L'analyse distributionnelle consiste à classer les éléments en fonction de leur distribution dans des combinaisons avec d'autres éléments. La principale difficulté de l'analyse distributionnelle est que tout est interdépendant et que les éléments servent à se classifier les uns les autres. Il existe ainsi une véritable symétrie entre classes distributionnelles et familles d'environnement. Ce type

de symétrie est appelé en mathématique la DUALITÉ. Le DUAL d'une classe distributionnelle est la famille des environnements compatibles avec les éléments de la classe et le DUAL d'une famille d'environnements est la classe distributionnelle des éléments compatibles avec les environnements de la famille. Autrement dit, si A est une classe distributionnelle, dual(A) = famille(A) = "l'ensemble des environnement compatibles avec A" et, si A est une famille d'environnements, dual(A) = classe(A) = "l'ensemble des signes compatibles avec A".

Le schéma suivant illustre la dualité. La dualité lie des OBJETS à des PROPRIÉTÉS : dans notre cas, les objets sont les signes et les propriétés les environnements (plus exactement la propriété est la compatibilité avec un environnement). Les propriétés définissent des classes d'objets, tandis que les objets définissent des familles de propriétés. Mais le procédé est totalement symétrique et on peut très bien considérer les propriétés comme nos objets et les objets comme leurs propriétés : dans ce cas, les « propriétés » d'une propriété donnée sont les objets compatibles avec elle.

Si A est une classe distributionnelle ou une famille d'environnements, dual(dual(A)) = A. La fonction « dual » associe donc une classe à chaque famille, laquelle est elle-même associée à cette classe. On en déduit qu'il y a exactement autant de classes distributionnelles que de familles d'environnements.

De plus, la fonction « dual » possède une excellente propriété qui est de **préserver la structure algébrique** définie par les opérations \cap et \cup et la relation d'ordre \subseteq , puisque dans les deux cas (que A et B soient des classes ou des environnements), on a :

- 1. si $A \subseteq B$, alors dual(B) \subseteq dual(A); autrement dit, plus on considère d'environnements, moins il y a de signes compatibles.
- 2. $dual(A \cup B) = dual(A) \cap dual(B)$; autrement dit, les signes compatibles avec à la fois les environnements dans A et ceux dans B sont, sans surprise, les signes qui sont à la fois compatibles avec A et compatibles avec B.
- 3. $dual(A \cap B) = dual(A) \cup dual(B)$.

La dualité inverse les rôles de \cap et \cup , ce qui ne change rien à la structure, car ces deux opérations sont elles-mêmes duales l'une de l'autre. L'ensemble des classes distributionnelles et l'ensemble des familles ont donc exactement la même structure algébrique et sont le **miroir l'un de l'autre**.

5.10 Signème

Nous nous sommes intéressés jusque-là au découpage des signes linguistiques selon l'AXE SYNTAGMATIQUE, c'est-à-dire selon la façon dont ils se combinent ou peuvent être décomposés. Nous allons maintenant nous intéresser à la délimitation des signes selon l'AXE PARADIGMATIQUE (voir la section 4.8 sur *L'identification des unités de la langue* et l'encadré ?? sur *La quatrième proportionnelle*). Nous regardons ici le paradigme des environnements pour les différentes occurrences d'une même forme afin de déterminer s'il s'agit des signifiants d'un même signe ou pas.

Considérons par exemple le segment /avãs-/ de *Le chat avançait*. On va retrouver le même segment /avãs-/ dans des énoncés tel que *Nous avançons grâce au vent* et *L'âne ne veut plus avancer*. On peut considérer qu'il s'agit du même signe, car sa forme est identique (l'alternance orthographique *c* vs *ç* n'est pas pertinente, la forme orale est identique) et que sa contribution sémantique est identique : 'avancer' signifie ici 'se déplacer vers l'avant'.

Considérons maintenant d'autres occurrences de /avãs-/ dans d'autres environnements :

Figure 5.2 – Découpage de /avãs-/ selon les axes paradigmatique et syntagmatique

Dans les énoncés *Ma montre avance* ou *Les travaux n'avancent pas très vite*, nous n'avons plus affaire au même signe /avãs-/, car le sens n'est plus 'se déplacer vers l'avant'. On peut néanmoins considérer qu'il s'agit, d'un certain point de vue, du même segment, car même si le sens est différent, il reste lié au premier sens que nous avons considéré. Il s'agit d'acceptions métaphoriques : lorsque ma montre avance, c'est comme si l'aiguille avançait plus vite que le temps; lorsque les travaux avancent, ils se rapprochent de leur achèvement. De plus, la forme du signe est exactement la même et sa combinatoire assez similaire : avanç- se combine toujours avec -ait (*Ma montre avançait*), mais par contre elle n'accepte plus de complément locatif (*L'âne avançait vers moi; *Ma montre avance vers moi*). Bien qu'il s'agisse de trois signes différents (puisque les signifiés sont différents), nous considérons pourtant qu'ils appartiennent à une même unité, qu'on nomme généralement le verbe AVANCER.

On retrouve encore le même segment /avãs/ dans des énoncés tels que *Pierre* est en avance, *Pierre a reçu une avance de 1000 euros, Pierre a fait des avances* à Marie. Ici le sens de /avãs/ est différent, mais surtout sa combinatoire est très différente : ce n'est plus un verbe et il ne peut plus se combiner avec un segment

tel que -ait. La forme est néanmoins identique et la parenté de sens non négligeable. Dans L'avancement des travaux est moins rapide que prévu ou Pierre a eu de l'avancement, /avãs/ est encore un signe (avancement est à avancer ce que changement est à changer), mais sa combinatoire est encore différente puisqu'il est indissociable de -ment.

Tout locuteur du français voit un lien entre toutes les occurrences d'/avãs/ que nous avons considérées. Nous considérons donc que, d'un certain point de vue (et d'un certain point de vue seulement), il s'agit toujours du même objet : un tel objet est appelé un signème.

Un signème est un ensemble de signes ou de quasi-signes de même forme et de sens apparentés. (Nous considérerons des signèmes de signes de formes différentes à la section 5.20 sur l'*Allomorphie*.)

On peut repérer à l'intérieur d'un signème des sous-ensembles de signes qui ont une distribution comparable. À l'intérieur du signème /avãs/, on repère ainsi trois (sous-)signèmes : le verbe AVANCER, le nom AVANCE et le radical *avance*- de AVANCEMENT.

5.11 Racine et signème **Q**

Jusqu'où pousser la recherche des segments communs? Il y a dans les mots commerce et marchand des segments comparables (/mɛrs/ pour commerce et /marʃ/ pour marchand) qui découlent de fait de la même RACINE indo-européenne /mɛrk/, que l'on retrouve non altérée dans mercantile. Néanmoins, il est probable que les locuteurs ordinaires du français (c'est-à-dire qui ne sont pas entraînés à analyser leur langue) n'auront pas remarqué cela, malgré la forte parenté sémantique de ces mots.

Qu'est-ce qui distingue racines et morphèmes?

Si nous reconnaissons dans *avancement* le même signème /avãs/ que dans *avançons*, c'est parce que *avancement* est à *avançons* ce que *changement* est à *changeons*. Rien de tel pour *commerce* et *marchand* où il n'y a aucune paire analogue. La mise en évidence d'un segment commun ne relève pas de la grammaire du locuteur et de sa connaissance **de** la langue, mais d'une connaissance **sur** la

langue. Le **signème** appartient à l'étude SYNCHRONIQUE de la langue, la **racine** à l'étude diachronique de la langue et à l'étymologie (l'étude de l'origine des mots). L'opposition entre SYNCHRONIE et DIACHRONIE a été posée par Ferdinand de Saussure; la synchronie considère un état de langue à un moment donné, tandis que la diachronie étudie les variations (du grec *dia*-'à travers') entre différents état de langue selon l'axe temporel (*chronos*).

Donnons deux autres exemples: avance, avant et avantage ont aussi une racine commune, comme le suggère les proximités de forme et de sens, mais on ne peut parler d'un même signème, puisqu'aucun des liens qui unissent ces mots n'a d'analogue en français. État et constater ont également une racine commune : la parenté de sens apparaît quand on pense à la synonymie entre constater et faire état et la parenté de forme devient évidente quand on pense qu'un autre sens de état se dit en anglais state (et que l'on a noté qu'il existe d'autres paires en français où un /s/ s'est effacé: été-estival, hôpital-hospitalier).

5.12 Signèmes minimaux : morphème et syntaxème

Si nous considérons maintenant le signème /avãs/, nous voyons qu'il n'est possible de décomposer aucun de ses éléments en deux signes linguistiques. On peut éventuellement essayer avec un segment /avã/ que l'on trouve dans *La plage est avant le village*, mais on ne voit pas comment la composition du sens de cet /avã/ avec un éventuel signe /s/ pourrait donner le signe *avanç*-/avãs-/. Proposer un signe /s/ n'aurait de sens qu'au cas où pour un autre signe X au moins, l'ajout d'un /s/ créait le sens 'se déplacer dans la direction X', c'est à dire si /avã/ et /s/ se combinaient proprement. Le signème /avãs/ est donc indécomposable.

Un MORPHÈME est un signème dont aucun signe n'est décomposable proprement.

Les morphèmes sont donc les signèmes minimaux du point de vue de l'opération de combinaison propre \boxplus . Nous pouvons considérer de même les signèmes minimaux du point de vue de la combinaison libre \oplus , que nous appelons les syntaxèmes.

Un syntaxème est un signème de signes de la même classe distributionnelle dont certains se combinent librement et dont aucun n'est une combinaison libre de signes.

Les morphèmes sont les unités minimales de la morphologie et les syntaxèmes sont les unités minimales de la syntaxe. La définition des syntaxèmes sera précisée dans le Chapitre 7.

Reprenons l'exemple du signème /avas/ : il s'agit d'un morphème, puisqu'aucun de ses éléments n'est décomposable proprement. Il contient deux syntaxèmes que sont avancer et avance. Enfin, avancement est un syntaxème, puisqu'aucune de ses acceptions n'est décomposable librement (seul un petit nombre de morphèmes verbaux permettent la combinaison avec -ment). Il est néanmoins décomposable proprement et contient une occurrence du morphème /avãs/.

5.13 Les termes morphème et autres X-èmes



La première étude des règles morphologiques d'une langue remonte à 2500 ans au moins. Il s'agit de l'extraordinaire description du sanskrit védique faite par le linguiste indien Pānini, qui comprend une description des morphèmes du sanskrit et des règles de combinaison de leurs signifiants. Les termes morphème et phonème sont introduits dans les années 1880 par le linguiste polonais Baudouin de Courtenay. La définition formelle des morphèmes et le principe de commutation doivent beaucoup aux travaux des distributionnalistes américains sur les langues amérindiennes (voir l'encadré ?? sur Structuralisme et distributionnalisme).

Les termes en -ème vont foisonner au cours du 20^e siècle. Les termes lexème et grammème sont couramment utilisés pour désigner respectivement les unités du lexique et celles de la grammaire. Nous introduisons syntaxème pour désigner les unités minimales de la syntaxe, qu'elles soient lexicales comme le *lexème* ou grammaticales comme le grammème. Le terme syntaxème (ou la variante syntactème) a été utilisé marginalement par quelques auteurs, mais ne s'est jamais imposé, d'autant que c'est le mot qui est généralement considéré comme l'unité minimale de la syntaxe (contrairement au point de vue défendu ici qui considère qu'il s'agit du syntaxème). Le terme sémantème, que nous utiliserons pour désigner les unités minimales de la sémantique, est beaucoup plus courant et a été utilisé par Charles Bally ou Igor Mel'čuk pour désigner les sens lexicaux. Nous

lui donnons un sens un peu différent en l'utilisant pour désigner des signes (et pas seulement des signifiés) (voir Chapitre 6 sur Sémantèmes et syntaxèmes). Tous les termes que nous venons de mentionner (morphème, lexème, grammème, syntaxème, sémantème) désignent, dans l'acception dans laquelle nous les utilisons, des ensembles de signes. Pour finir, nous introduisons le terme signème pour désigner tout ensemble de signes d'un de ces types.

5.14 Lexème, flexion et grammème

On distingue deux principaux types de syntaxèmes : les syntaxèmes lexicaux ou lexèmes et les syntaxèmes flexionnels ou grammèmes.

- AVANCER est un SYNTAXÈME LEXICAL ou LEXÈME : il appartient à une classe distributionnelle ouverte de syntaxèmes, c'est-à-dire qu'il peut commuter avec un très grand nombre de syntaxèmes, potentiellement illimité (cet ensemble est celui des verbes). De plus, ses sens sont assez précis pour être paraphrasés. Par exemple, dans l'énoncé Le chat avançait, AVANCER signifie 'se déplacer vers l'avant' et il commute avec sa définition : Le chat se déplaçait vers l'avant est une paraphrase de Le chat avançait.
- -ait est une flexion. Il s'agit d'une combinaison de plusieurs syntaxèmes, qui comprend notamment la combinaison libre d'un temps (l'imparfait) et d'un accord en nombre et personne. Les syntaxèmes qui composent une flexion s'appellent des syntaxèmes flexionnels ou grammèmes. Flexions et syntaxèmes flexionnels appartiennent à des classes distributionnelles fermées d'éléments, c'est-à-dire qu'ils commutent avec un nombre restreint d'éléments similaires dont on peut faire la liste exhaustive (il y a par exemple 51 flexions possibles pour un verbe du français : 6 formes pour chacun des 7 « temps » simples, 4 formes pour chacun des deux participes, auxquelles il faut encore ajouter la forme infinitive; 51 = 7×6 + 2×4 + 1). Les grammèmes expriment des significations grammaticales qui ne peuvent être paraphrasées facilement.

Lexème verbal et flexion ne peuvent donc pas être utilisés de manière Autonome. Le lexème et sa flexion sont indissociables l'un de l'autre : une flexion verbale ne s'utilise pas sans une base verbale et un lexème verbal ne s'utilise pas sans flexion. Ceci nous amène à la définition du mot, sur laquelle nous reviendrons au Chapitre ?? : un mot est grosso modo un signe linguistique autonomisable minimal, c'est-dire dont les parties sont indissociables et donc non-autonomisables.

Du fait que les lexèmes verbaux ne sont pas autonomisables, il est d'usage de les nommer par l'une de leurs formes fléchies. L'usage en français est d'utiliser la forme infinitive. Cet usage est purement conventionnel (et pédagogiquement assez mauvais puisque le radical de l'infinitif n'est généralement pas le radical de base; voir l'encadré ?? sur *Syntaxème zéro et troncation en français*); par exemple en grammaire latine, il est d'usage de nommer un lexème verbal par la forme de la 1^{ère} personne du singulier du présent (lat. Amo 'j'aime') et en grammaire arabe par la forme de la 3^{ème} personne du singulier du passé (ar. KATABA 'il a écrit'). Pour éviter toute confusion entre le lexème et la forme infinitive, nous notons le premier en petites majuscules, AVANCER, et la deuxième en italique, *avancer*. Les syntaxèmes flexionnels sont désignés par des termes métalinguistiques : présent, pluriel, etc. La forme *avancer* est la combinaison de AVANCER et de l'infinitif :

(6) $avancer = AVANCER \oplus infinitif.$

5.15 Signème libérable, radical, affixe

Nous allons caractériser les occurrences des morphèmes au sein des lexèmes, c'est-à-dire les MORPHÈMES SOUS-LEXICAUX. Un lexème qui peut être décomposé en plusieurs morphèmes est dit COMPLEXE.

Un signème est dit LIBÉRABLE s'il existe des environnements dans lesquels ce signème commute librement. Un MORPHÈME LIBÉRABLE est donc un morphème qui contient un syntaxème. Lorsque ce syntaxème est lexical, le morphème est dit LEXICAL.

Cette propriété est utilisée pour classifier les morphèmes constitutifs d'un lexème complexe. Si l'on considère le syntaxème *brûlure*, qui se décompose en *brûlmure*, on constate que *brûl*- est libérable, puisqu'il commute librement dans *nous brûlons*, mais -*ure* n'est pas libérable, puisqu'il n'existe aucun environnement où -*ure* commute librement.

Un morphème, lorsqu'il est l'unique morphème lexical d'un lexème, est appelé le RADICAL du lexème. Un morphème non lexical est appelé un AF-

FIXE. Lorsqu'il précède le radical, il s'agit d'un PRÉFIXE et lorsqu'il le suit d'un suffixe.

Il existe également des lexèmes construits uniquement avec des morphèmes non libérables. En français, un certain nombre de lexèmes, appelés composés savants, sont construits avec deux morphèmes lexicaux empruntés au grec ancien (dont certains sont devenus libérables) : sismo⊞graphe, sismo⊞logue, grapho⊞logue, géo⊞graphe, géo⊞thermie, thermo⊞mètre, métro⊞nome, etc. De tels morphèmes, dont le statut est intermédiaire entre affixe et radical, sont appelés des CONFIXES, car ils doivent être associés par paire pour former un lexème.

Il faut noter que les termes de *libre* et de *libérable* sont définis par des critères purement distributionnels et recouvrent donc des notions syntaxiques. Ils ne doivent pas être confondus avec un autre emploi du terme *libre* dû à Bloomfield qui est proche de ce que nous avons appelé l'autonomisabilité et qui conduit à appeler *forme liée* tout affixe. Dans notre terminologie, au contraire, les **syntaxèmes flexionnels**, bien qu'étant des affixes, se combinent librement par définition et sont parfaitement *libérables*. Par contre, ils ne sont **pas autonomisables**, puisqu'ils sont indissociables de lexèmes.

Les notions de radical et d'affixe doivent être étendues par analogie. Ainsi, un quasi-morphème comme *struct*-, bien que non libérable, est également considéré comme un élément lexical et comme le radical de *structure* et de *construction*, car il commute avec des morphèmes lexicaux.

A l'inverse, certains préfixes du français présentent un cas limite d'affixe, puisqu'ils appartiennent à des morphèmes prépositionnels comme le *en* de *il enterre* ou le *sur* de *il surestime*. On considère quand même qu'il s'agit d'affixes, car les prépositions simples forment une classe fermée et sont donc moins lexicales que les morphèmes qui ont des emplois en tant que syntaxèmes appartenant à des classes ouvertes. De plus, la classe distributionnelle des préfixes du français contient quand même une bonne proportion de morphèmes non libérables, comme *con*- ou *in*-, et qui commutent avec les autres (*comprendre* vs *sur prendre*, *inestimable* vs *sur estimer*).

5.16 Dérivation et composition

La plupart des lexèmes complexes du français sont construits avec un radical et un certain nombre d'affixes. Par exemple, le lexème *détournement* est construit

avec le radical tourn-, le préfixe dé- et le suffixe -ment. Les lexèmes complexes comportant un unique morphème lexical sont appelés des dérivés morpholo-GIQUES. L'opération qui consiste à ajouter un affixe à un lexème pour former un nouveau lexème est appelée la DÉRIVATION MORPHOLOGIQUE.

Il existe aussi des lexèmes construits avec plusieurs morphèmes lexicaux, comme soutien-gorge ou bonhomme : de tels syntaxèmes sont appelés des composés mor-PHOLOGIQUES. L'opération qui consiste à combiner deux lexèmes pour former un nouveau syntaxème est appelée la composition Morphologique. Un composé peut être la source d'une dérivation comme dans bonhommie.

On peut associer à un lexème complexe une structure morphologique en regardant les portions du lexème qui sont libérables. Une UNITÉ MORPHOLOGIQUE est soit un morphème, soit une portion libérable d'un lexème. Les unités morphologiques de détournement sont les morphèmes dé-, tourn- et - ment et les combinaisons détourn- (il détournait) et détournement. Par contre °tournement n'est pas libérable. On en déduit une STRUCTURE MORPHOLOGIQUE que nous pouvons représenter dans ce cas par un parenthésage : détournement = (dé⊞tourn)⊞ment. Ceci induit un CHEMIN DÉRIVATIONNEL menant du radical du lexème au lexème complet par combinaisons successives de morphèmes. Pour détournement, le chemin est $tourn(er) \rightarrow d\acute{e}tourn(er) \rightarrow d\acute{e}tournement$ et pas $tourn(er) \rightarrow {}^{\circ}tournement$ → détournement. Cela signifie que, dans la façon dont est perçu le lexème détournement, c'est dé- qui s'affixe d'abord à tourn-, puis -ment au tout.

L'étude des morphèmes sous-lexicaux, c'est-à-dire des occurrences d'un morphème à l'intérieur d'un lexème, ne relève pas de la syntaxe et est donc hors de la visée de cet ouvrage. Une telle étude concerne la construction des lexèmes, c'està-dire la morphologie constructionnelle et la lexicologie. Cette notion est néanmoins essentielle à la compréhension de ce que sont les syntaxèmes et de ce qui les différencie des occurrences liées des morphèmes.

5.17 « Syntaxe » des morphèmes sous-lexicaux 🝳



Lorsqu'on parle de syntaxe, on parle normalement de combinatoire libre (voir la section 7.6 sur Syntaxe et morphologie) et lorsqu'on parle de combinatoire libre, on parle de combinaisons qui sont faites par le locuteur au moment de la production d'un nouvel énoncé (voir l'encadré ?? sur Liberté de combinaison et opposition parole/langue). La combinatoire des sous-morphèmes relève au contraire de la structure d'unités déjà construites et stockées dans le lexique mental du locuteur (même s'il existe la possibilité de construire de nouvelles unités lors de la production avec les constructions productives). Il en découle que l'histoire



constructionnelle de ces unités se fait sur un temps beaucoup plus long, bien qu'il en reste des traces en synchronie (par la commutation propre avec d'autres morphèmes). On peut alors se poser la question suivante : la structure des combinaisons de sous-morphèmes est-elle ou non de la même nature que celle des syntaxèmes?

Nous étudierons en détail la structure des combinaisons syntaxiques dans la partie 3, dont c'est le sujet central. Les outils que nous introduirons peuvent aussi s'appliquer à la description des combinaisons morphologiques. Nous opposerons en particulier deux approches principales de la structure syntaxique, la structure de dépendance et la structure de constituants. La principale différence entre les deux approches reposent sur ce qu'elles encodent prioritairement (voir l'encadré ?? Dépendance et constituance se complètent). Les structures de constituants permettent d'encoder naturellement l'ordre dans lequel les combinaisons ont lieu et c'est ce que nous avons utilisé pour présenter l'analyse de détournement, car nous voulions indiquer que lors de l'histoire dérivationnelle de ce mot, tourns'est d'abord combiné avec dé-, puis -ment. Ils existent des cas où l'histoire dérivationnelle n'est pas aussi claire en synchronie. Tel est le cas par exemple de châtaigneraie, pommeraie ou orangeraie. Il s'agit a priori du suffixe -aie qui appliqué à un nom d'arbre désigne un groupe d'arbre, comme dans chênaie 'groupe de chênes' ou peupleraie 'groupe de peupliers', avec en plus une alternance -ier /je/ \rightarrow -er/k/. On aurait donc châtaigneraie = (châtaigne \boxplus ier) \boxplus aie. Mais on peut aussi analyser ces dérivations comme le résultat direct de l'application du suffixe -eraie, lequel suffixe se rencontre dans chêneraie 'groupe de chênes' (lexème considéré comme fautif mais largement attesté). On peut représenter cette ambiguïté en utilisant une représentation non parenthésée de la dérivation (châtaigneraie = châtaigne⊞ier⊞aie), qui s'apparente alors à une structure de dépendance, puisqu'on n'indique que -ier se combine aussi bien avec châtaigne que -aie sans spécifier quelle combinaison a lieu en premier.

5.18 Sémantique des morphèmes sous-lexicaux



Le fait qu'un affixe soit toujours un morphème non libérable influe sur la nature même de son sens : l'affixe ne peut jamais être « isolé » et son sens ne peut pas être saisi de manière autonome. Le sens d'un affixe n'apparaît que dans la relation entre deux lexèmes : le sens de -ure naît de la relation entre BLESSURE et BLESSER. Ainsi le sens d'un tel morphème est nécessairement de nature opératoire : il n'est exprimable que dans la relation entre deux sens. Le sens de -ure est le sens de l'opération qui permet de construire 'blessure' à partir de 'blesser' **(1)**

et d'autres paires du même type : une X-ure est le résultat obtenu en X-ant.

Il existe aussi des morphèmes, comme le -aume de royaume, qui n'apparaissent que dans un seul lexème dont la diagrammaticité est suffisamment claire pour qu'on attribue un sens au morphème : le fait que le lien sémantique entre royaume et roi soit assez simple (un royaume est le territoire sous la tutelle d'un roi) et que la paire roi-royaume soit parallèle à d'autres paires comme prince-principauté ou duc-duché, permet de définir facilement le sens (opératoire) de -aume. Par contre, fur, qui est pourtant un mot, mais qui n'apparaît que dans au fur et à mesure n'a plus de sens accessible en français contemporain, car la construction dans laquelle il entre est isolée et qu'il n'y a jamais de commutation possible sur fur. De même, beaucoup de locuteurs du français vont utiliser une locution comme être dans le collimateur sans avoir aucune idée de ce qu'est un collimateur (voir la section ?? sur Le collimateur et la sellette pour d'autres exemples).

5.19 À la limite entre morphème sous-lexical et syntaxème flexionnel

La notion de commutation libre est GRADUELLE et il est parfois difficile pour certains morphèmes de les situer entre morphème sous-lexical et syntaxème (flexionnel). Le cas du morphème -ment qui transforme un adjectif en adverbe est un cas intéressant de morphème qui ne commute pas librement, mais s'en approche néanmoins. Dans de nombreux cas, le changement de partie du discours opéré par -ment ne s'accompagne d'aucun changement de sens, ce qui signifie que la commutation est propre : cf. le passage de un départ rapide à partir rapidement ou de une réflexion intense à réfléchir intensément. Mais la dérivation en -ment n'est pas systématiquement possible : cf. une réflexion poussée/élaborée vs *réfléchir poussément/élaborément, un départ inattendu vs *partir inattendument. Même des adjectifs très courant comme gros n'ont pas d'adverbe correspondant : faire une grosse erreur vs *se tromper grossement. Certains adverbes en -ment sont fort peu diagrammatiques et ne peuvent pas être décomposés proprement, comme vraiment, carrément ou vertement. En conséquence, la classe des éléments qui se combine avec -ment comporte pas mal d'irrégularités et est par exemple différente de la classe des adjectifs qui se combinent avec de façon, puisqu'on a de façon poussée, élaborée ou inattendue et pas de façon verte. On peut également noter quelques irrégularités morphologiques (voir la section 5.20 sur l'Allomorphie) dans la combinaison de -ment et d'un adjectif, toutefois la combinaison libre des verbes avec leur flexion comprend davantage encore d'idiosyncrasies, donc les arguments purement morphologiques ne permettent pas de décider s'il s'agit de



combinaison libre ou non.

Un autre cas discutable est le « genre » des noms. Le genre des noms ne doit pas être confondu avec le genre des adjectifs. Le genre des adjectifs est un syntaxème flexionnel qui se combine librement avec l'adjectif. Il n'a aucune contribution sémantique et sert seulement à mieux marquer la dépendance entre le nom et l'adjectif qui le qualifie (ce rôle syntaxique constitue la signification du syntaxème flexionnel). Certains noms varient apparemment en genre comme lion vs lionne, acteur vs actrice ou prince vs princesse. Mais dans ce cas, l'opposition est sémantique et marque une opposition de sexe : la lionne est la femelle du lion. Elle peut s'accompagner d'une variation de sens plus importante, puisque par exemple un prince est toujours le fils d'un roi, tandis qu'une princesse peut être la fille d'un roi ou la femme d'un prince. De même, un lion et une lionne possèdent des différences morphologiques suffisamment importantes, comme la crinière, pour qu'il soit bizarre de dire qu'une lionne est un lion femelle. De plus, cette alternance de genre est assez capricieuse avec de nombreux noms dons le genre est fixe (grenouille, girafe, chacal, moustique, etc.), même si l'alternance de sexe reste signifiante et plusieurs cas où deux formes sans liens morphologiques existent (poule et coq ou jument et étalon). Citons encore les nombreux trous que des néologismes comme auteure ou professeure peinent encore à combler. Il y a donc toutes les raisons de considérer le /ɛs/ de /preses/ princesse comme un morphème sous-lexical plutôt qu'un syntaxème flexionnel.

5.20 Allomorphie

Nous n'avons jusque-là considéré que des signèmes dont tous les éléments ont le même signifiant. Il y a néanmoins de bonnes raisons d'élargir la notion de signème et d'accepter des alternances de formes. Ainsi là où le verbe AVANCER présente une seule forme /avãs-/, le verbe ALLER présentent quatre formes : /v/dans tu vas, /al/- dans nous allons, /i/- dans nous irons, /aj/- dans qu'il aille. On veut pourtant regrouper ces formes parce que all+ons est à avanç+ons ce que i+rons est à avance+rons. Nous dirons que all- et i- sont des allomorphes.

Pour que les signes A_1 et A_2 soient des allomorphes, il faut que A_1 et A_2 aient des signifiants différents, mais qu'il existe un signe A' et un certain nombre de signes B et B' tels que A_1 +B soit à A_2 +B' ce que A'+B est à

A'+B'.

Cette propriété est bien vérifiée par notre premier exemple : $A_1 = al$ -, $A_2 = i$ -, A' = avanç-, B = -ons, B' = -rons et $A_1 + B = allons$ est à $A_2 + B' = irons$ ce que A' + B = avançons est à A' + B' = avancerons. Deux allomorphes sont donc des signes différents, puisqu'ils ont des signifiants différents, mais ils se comportent ensemble comme s'ils formaient un unique signe, puisqu'ils commutent avec un même signe A'. On traduit cela en demandant qu'ils commutent proprement avec un tel signe dans un certain nombre d'environnements.

La notion d'allomorphie nous permet d'étendre la notion de morphème.

Un MORPHÈME POLYMORPHIQUE est la réunion de plusieurs « morphèmes » de formes différentes qui sont des allomorphes les uns des autres. Un MORPHE est un sous-ensemble de signes d'un morphème qui possèdent le même signifiant.

La première propriété que nous avons donnée pour définir l'allomorphie n'est pas suffisante. Une autre propriété nécessaire pour être des allomorphes est de se ressembler. Il y a en fait deux types de ressemblance possibles et ainsi deux notions d'allomorphie : l'une, syntaxique, est une allomorphie au sein d'un syntaxème, basée sur une similarité de syntactique ; l'autre, morphologique, est une allomorphie au sein d'un morphème, basée sur la similarité des formes.

Les signes A_1 et A_2 sont des allomorphes (syntaxiques) si A_1 et A_2 ont des formes différentes, mais des sens identiques et s'il existe un syntaxème A' dont l'ensemble des contextes possibles réunit les contextes possibles de A_1 et A_2 .

Le cas de Aller vu précédemment illustre l'allomorphie syntaxique. La réunion des paradigmes de flexion des différents allomorphes de Aller comprend 51 formes et est équivalent au paradigme de flexion de n'importe quel verbe.

Un autre exemple d'allomorphie syntaxique est l'alternance des formes $/\epsilon/$ vs /j/ pour l'imparfait. En effet, ces deux formes ont exactement les mêmes sens et elles possèdent, à elles deux, la même extension que le syntaxème zéro du présent ou le syntaxème / ϵ / du futur.

Le deuxième cas d'allormorphie nous concerne moins, car il ne relève pas de la syntaxe.

Les signes A_1 et A_2 sont des allomorphes (morphologiques) si A_1 et A_2 ont des formes différentes mais similaires, des sens apparentés et s'il existe des signes A', B et B' tels que A_1 +B soit à A_2 +B' ce que A'+B est à A'+B'.

Dans le cas de l'allomorphie morphologique, il n'y a plus identité de sens et la combinatoire est généralement très différente : il faut donc que A_1 et A_2 possèdent une proximité formelle suffisante pour qu'on considère qu'ils sont parents. Par exemple, écriv- et écrit- sont des allomorphes, car écrit+ure est à écriv+ons ce que sculpt+ure est à sculpt+ons et /ekʁit/ et /ekʁiv/ sont formellement proche. La différence de forme peut être plus importante si elle se retrouve plusieurs fois : construct- et construis- sont des allomorphes, car construction est à construisons ce que instruction est à instruisons ou ce que destruction est à détruisons.

Par contre, même si *coup* est à *frapp+ons* ce que *gifle* est à *gifl+ons*, on ne dira pas que *coup* est un allomorphe de *frapp-*. Il y a là une trop grande distance formelle. On dira plutôt que *coup* est un SUPPLÉTIF de *frapp-*.

En général, les allomorphes d'un signème sont en distribution complémen-

TAIRE, c'est-à-dire que leurs environnements respectifs sont incompatibles (voir la notion de *complémentaire* en *Théorie des ensembles* dans la section ??). Mais il existe aussi des cas d'alternance de forme dans un même environnement. Par exemple le verbe BALAYER possèdent deux allomorphes /balɛ/ et /balɛj/ qui peuvent alterner avec certaines flexions : *ils balaient* vs *ils balayent*.

5.21 Alternance vocalique en français : allomorphie ou alternance phonologique?

Identifier le nombre d'allomorphes d'un morphème ne va pas nécessairement de soi. Ainsi le verbe céder alterne apparemment deux formes : céd- [sed] dans nous cédons et cèd- [sɛd] dans il cède. Si [e] et [ɛ] peuvent être distinctifs dans les syllabes ouvertes (CV) finales (cf. l'opposition entre été /ete/ vs était /ete/ en français parisien), ils se neutralisent dans les autres positions et forment ce qu'on appelle un archiphonème, que nous noterons /E/ : en syllabe fermée, /E/ est prononcé [ɛ] et, en syllabe ouverte non finale, il est prononcé /e/. Ainsi l'alternance relève complètement de la phonologie du français et CÉDER possède donc un unique morphe /cEd/. La situation est différente pour APPELER qui présente une ALTERNANCE entre appell- /apɛl/ dans il appelle et appel- /apəl/ dans nous appelons: ici l'alternance n'a plus de motivation phonologique en français contemporain et il y a bien allomorphie. Le fait que la distribution des deux allomorphes de Appeler soit identique à celle des deux prononciations de Céder laisse supposer que cette allomorphie résulte d'une alternance phonologique aujourd'hui morte, du même type que celle bien vivante de CÉDER. Une allomorphie de même distribution (et de même origine) se trouve avec les verbes MOURRIR ou POUVOIR (il meurt vs nous mourons, il peut vs nous pouvons). Nous verrons un autre cas distributionnellemnt similaire à l'allomorphie avec les verbes dit du 2ème groupe (dans l'encadré ?? sur Syntaxème zéro et troncation en français).

5.22 Amalgame, alternance et mégamorphe

Nous avons fait jusque-là comme si les signifiants se combinaient de manière CONCATÉNATIVE, c'est-à-dire en s'enchaînant les uns à la suite des autres. Il existe certaines situations où il est difficile de trouver la limite entre un lexème et sa flexion. Par exemple, la forme verbale [nous] sommes entre dans le paradigme des formes du verbe ÊTRE et se décompose normalement en un morphe du morphème ÊTRE et une flexion. Les formes en -mes de la 1ère personne du pluriel ne

se rencontrent plus qu'au passé simple ($nous \, chant + \hat{a} + mes$) et de surcroît aucune autre forme de $\hat{E}TRE$ n'a le radical som-. Il s'agit donc, pour sommes, d'une forme amalgamée exprimant conjointement le lexème $\hat{E}TRE$ et sa flexion.

Une combinaison de syntaxèmes dont le signifiant est INDÉCOMPOSABLE est appelée un AMALGAME ou MÉGAMORPHE FORT. Les syntaxèmes au sein d'un amalgame sont réalisés de manière CUMULATIVE.

Même des syntaxèmes normalement réalisés par des mots séparés peuvent s'amalgamer, comme *au* /o/ en distribution complémentaire avec à *la* /ala/. Les flexions, c'est-à-dire les combinaisons de syntaxèmes flexionnels, sont souvent des mégamorphes forts: par exemple, si on considère que l'accord verbal -ons /ɔ/ combine un accord en personne (1ère personne) et un accord en nombre (pluriel) (voir discussion sur ce point dans la section ?? sur les *Catégories flexionnelles abstraites*), cette combinaison est un mégamorphe fort.

Un signe tel que *chevaux* /ʃəvo/ pose un problème un peu différent. Ici le principe de commutation s'applique à l'ensemble du signe, signifiant compris : *chevaux* est à *cheval* ce que *animaux* est à *animal*. Il n'est néanmoins pas satisfaisant de décomposer *chevaux* en *chev+aux*, car les dérivés de *cheval/chevaux* prennent *cheval* comme base : chevalin, chevalier ou *chevaleresque*. Même chose pour les dérivés d'autres lexèmes en -al/-aux : *animal* ité, *métallique*, *canaliser*, etc. Il faut donc considérer que le radical du lexème cheval est *cheval* et que le pluriel est réalisé par une Alternance /al/ \Rightarrow /o/. Le signifiant de cet allomorphe du pluriel est ainsi traité comme une opération qui transforme le segment terminal du radical /al/ en /o/. En conséquence, le signifiant du pluriel ne peut donc pas être séparé du signifiant lexical dans la forme /ʃəvo/ et cette forme est ainsi non segmentable, bien qu'elle soit obtenue par la **combinaison régulière** du signifiant lexical et d'une apophonie.

Un signe dont le signifiant est décomposable sans être segmentable est appelé un MÉGAMORPHE FAIBLE.

5.23 Flexions irrégulières 🍳

Pourquoi les formes verbales du présent sont plus irrégulières que celle des autres temps? Pourquoi les participes passés sont plus irréguliers que les participes présents? Pourquoi les formes des verbes très utilisées comme être, avoir ou aller sont plus irrégulières que les autres? Parce que les formes les moins courantes sont construites par analogie par les locuteurs et sont donc régulières, alors que les formes très courantes sont apprises dans leur globalité. Dans ces formes qui ne sont plus décomposées (ou plutôt qui ne sont pas composées au moment de leur utilisation), les signifiés des syntaxèmes tendent à se fusionner. Ceci est vrai dans toutes langues flexionnelles : l'irrégularité la plus grande concerne toujours les mots les plus courants. Diachroniquement, les hausses et baisses de fréquences d'un mot sont souvent suivis d'une acquisition ou d'une perte d'irrégularité, ce qui est un argument en faveur de la fréquence d'usage en tant que variable dans la description linguistique.

5.24 Syntaxème zéro

Le cas des syntaxèmes zéros est bien illustré avec la conjugaison du français : lorsqu'on considère les quatre formes suivantes du verbe AVANCER à la 1ère personne du pluriel, [nous] avançons, avancerons, avancions, avancerions, on voit que le morphème d'accord -ons est obligatoire et que les morphèmes -er- et -i- peuvent s'intercaler et eux seuls. Dans nous avançons, l'absence de ces morphèmes indique que le procès exprimé par le verbe AVANCER a lieu au moment de l'énonciation (pour une description détaillée des syntaxèmes de temps du français on consultera la section ?? sur le Temps verbal). La forme avançons s'oppose aux autres formes — avancerons, avancions, avancerions — qui expriment nécessairement le fait que le procès n'a pas lieu au moment de l'énonciation et ont donc un sens incompatible avec celui de avançons. L'absence de -er- et -i- est donc signifiante.

Lorsque l'absence d'un syntaxème est signifiante, on peut postuler la présence d'un syntaxème zéro, c'est-à-dire un syntaxème dont la réalisation morphologique est un segment vide. Un syntaxème zéro est noté \emptyset dans les analyses morphologiques.

Les syntaxèmes zéros sont des objets abstraits introduits par les linguistes pour modéliser le caractère obligatoire d'un autre syntaxème. C'est parce que le pluriel doit obligatoirement être marqué sur les noms en anglais, la plupart du temps par -s, et que donc le singulier est marqué par l'absence de ce morphème, que l'on peut dire que le singulier des noms en anglais est marqué par un syntaxème zéro. Dans l'exemple suivant, la première ligne donne le mot anglais décomposé au niveau morphologique et la deuxième ligne sa GLOSE, indiquant les syntaxèmes en jeu :

(7) cat-Ø cat-s chat-SG chat-PL

Les syntaxèmes zéros appartiennent toujours à un paradigme fermé, condition sine qua non pour que l'absence soit signifiante. Ce sont donc nécessairement des syntaxèmes flexionnels.

On peut contraster le cas des syntaxèmes zéros avec celui, plus habituel, illustré par la paire *nous avançons* vs *nous avançons vite*. Ici l'absence de *vite* n'est pas signifiante et le sens 'nous avançons' n'est pas incompatible avec le sens 'nous avançons vite'. Le syntagme *nous avançons* est simplement moins spécifié que *nous avançons vite*, alors qu'on ne peut pas dire que *nous avançons* est moins spécifié que *nous avançons* ou *nous avancerons*.

Nous ne parlerons pas de *morphème zéro* comme c'est souvent l'usage. Nous allons même plus loin : la notion de *morphème zéro* n'existe pas, car un syntaxème zéro n'a pas de réalisation phonologique et morphologique. C'est justement l'absence de morphème qui est signifiante. Si le syntaxème zéro possède un signifié similaire à ceux des éléments avec lesquels il « commute », il ne possède pas de signifiant morphologique. On peut voir le syntaxème zéro comme un demi-signe dont le signifiant serait purement syntaxique, purement combinatoire (voir la section 6.18 sur *Syntaxème et faisceau de signes*, où la notion de demi-signe est développée). La notion de zéro est une notion purement syntaxique, qui met nécessairement en jeu une catégorie flexionnelle, c'est-à-dire un paradigme de signes qui commutent librement.

On fera attention à ne pas confondre syntaxème zéro et **syntaxème vide**. Nous utilisons le terme *vide* pour les signes sémantiquement vides, c'est-à-dire qui n'ont pas de contribution sémantique propre, ce qui est la situation inverse de celle des syntaxèmes zéros (voir la section 6.4 sur le *Syntaxème vide*). Néanmoins, nous reprenons pour les syntaxèmes zéros la notation traditionnelle \emptyset . Ce symbole, qui désigne l'ensemble vide en mathématique (voir l'encadré ?? sur la *Théorie des ensembles*), exprime bien le fait que la position « occupée » par le syntaxème zéro reste vide.

5.25 Syntaxème zéro et troncation en français 🛦

Accord en genre des adjectifs. L'accord des adjectifs en français à l'écrit présente un cas de syntaxème zéro. L'adjectif possède quatre formes : vert, verte, verts, vertes. On voit, sur ces formes écrites, que les deux segments -e- et -s peuvent s'ajouter au lexème vert-. Comme le fait d'ajouter ou de ne pas ajouter ces morphèmes est une décision obligatoire, l'absence de ces morphèmes est signifiante. On est donc conduit, pour l'analyse des formes écrites, à introduire deux syntaxèmes zéros et d'avoir ainsi un système avec deux oppositions : \emptyset_1 vs -e- correspondant à l'opposition masculin vs féminin, \emptyset_2 vs -s correspondant à l'opposition singulier vs pluriel.

L'étude des formes orales des adjectifs présente un paysage un peu plus complexe : en effet, si l'on regarde les formes /vɛʁ/ vs /vɛʁ/ (vert vs verte à l'écrit), on voit que le féminin fait apparaître le phonème /t/, lequel segment sonore, dans une application simple et simpliste du principe de commutation, pourrait être considéré comme le signifiant du féminin. De même, les formes /blã/ vs /blã[/, /gki/ vs /gkiz/, /gko/ vs /gkos/, /gkã/ vs /gkãd/ (blanc(he), gris(e), gros(se), grand(e)) devraient nous amener à considérer que le féminin est successivement exprimé par /ʃ/, /z/, /s/, /d/. Or les mêmes segments apparaissent dans les dérivés blancheur, grisaille, grosseur ou grandeur et font donc partie du signifiant du radical qui sert à les former. Il paraît donc plus judicieux de considérer que ces consonnes finales font partie du signifiant du lexème adjectival et que le masculin se réalise par une Troncation de la consonne finale du radical. On peut hésiter à considérer que la troncation fait partie du signifiant du syntaxème masculin ou bien qu'elle résulte d'une propriété du radical. Nous préférons la deuxième solution : la chute de la consonne finale du radical peut être vue comme due à l'absence d'un support pour cette consonne, ce qui nous amène à considérer que le masculin est exprimé par un syntaxème zéro et le féminin par quelque chose qui sert de support à la consonne et qu'on peut modéliser par un syntaxème de

signifiant /-ə/ (e muet), modélisation qui est à l'origine de nos conventions orthographiques.

Conjugaison des verbes. Un autre cas intéressant de troncation est celui des verbes du 2^{ème} groupe qui alternent deux formes : sali- dans il salit et salissdans nous salissons. Les dérivés, comme salissure ou polissage, sont construits sur la forme longue. La forme courte est utilisée pour trois formes du présent (je salis, tu salis, il salit), le participe passé (sali), l'infinitif (salir), le futur (nous salirons) et le conditionnel (nous salirions). On peut faire ici la même analyse que pour la flexion de l'adjectif (/νεκ/ vs /νεκt/) et considérer que, en l'absence d'une voyelle de soutien, la consonne finale du radical verbal tombe : les flexions singulier du présent et du participe passé ont donc un signifiant zéro, tandis celle de la 3^{ème} personne du pluriel du présent (ils salissent) ou celles du subjonctif (qu'il salisse) ont un signifiant -/ə/ qui maintiendra la consonne finale. Futur et conditionnel sont exprimés par un même morphème consonantique, -/ʁ/- qui ne permet pas non plus à la consonne finale du radical de se maintenir. On en conclut que le verbe SALIR possède un morphe unique, /sali(s)-/, dont la consonne finale chute en l'absence d'un voyelle de soutien. Comme le signifiant unique de SALIR n'est pas complètement prononcé dans la forme infinitive (salir), il serait pédagogiquement plus judicieux de présenter les verbes dans une forme où le lexème apparaît dans sa forme complète (par exemple à l'imparfait qui est la forme la plus régulière des verbes français : il salissait).

On retrouve la même conjugaison pour des verbes dit du 3ème groupe comme ÉCRIRE ou CONSTRUIRE : il écrit vs il écrivait, il construit vs il construisait, mais avec des consonnes finales du radical différentes (/v/ et /z/ au lieu de /s/) et une orthographe différente de l'infinitif peu justifiée (écrire vs salir). Ces verbes présentent néanmoins des différences de conjugaison au participe passé (la lettre est finie vs la lettre est écrite) et au passé simple (il salit vs il écrivit). D'autres verbes du français possèdent également une consonne finale qui chute devant un morphème zéro (il vit vs il vivait, il part vs il partait, il dort vs il dormait), mais qui se maintient devant le syntaxème du futur -/B/- (vivre, nous vivrons), éventuellement grâce à une voyelle épenthétique (partir, nous partirons).

5.26 Syntaxème et morphologie

Cet ouvrage s'intéresse peu aux morphèmes, c'est-à-dire aux unités minimales de forme. Nous nous intéressons à la combinatoire des unités et donc aux unités minimales du point de vue de la combinatoire (libre), ce qui nous a amené à introduire la notion de *syntaxème*. Les syntaxèmes tendent à être aussi des uni-

tés minimales de forme, notamment les syntaxèmes flexionnels, ce qui entraîne généralement une confusion entre les notions de *morphème* et de *syntaxème*, accentuée par le fait que la troisième composante du signe, le syntactique, n'est généralement pas considérée. Nous nous intéressons à la MORPHOLOGIE, c'est-à-dire à l'étude de la combinatoire des signifiants, lorsqu'il s'agit de la combinaison de syntaxèmes. Nous avons introduit dans ce chapitre plusieurs notions qui relèvent directement de la morphologie, comme l'amalgame, l'alternance et la troncation. Nous en reparlerons dans l'encadré ?? sur les *Langues isolantes, agglutinantes et flexionnelles*.

On notera pour conclure que nous n'avons pas encore parlé de la notion de *mot*. Celle-ci ne sera définie qu'à la section ?? intitulée *Mot*. L'ensemble de la partie 4 est consacrée à la NANOSYNTAXE, c'est-à-dire à l'étude des combinaisons de syntaxèmes particulièrement cohésives, appelée traditionnellement *morphologie flexionnelle*.



Exercices

Exercice 1 a) Montrer que la décomposition nation + al est propre. b) Pourquoi la décomposition $id\acute{e}(e) + al$ est-elle moins diagrammatique?

Exercice 2 a) En 2002, Michel Tournier a écrit un livre intitulé « *Journal extime* ». De quelle façon ce titre joue-t-il sur les morphèmes? Même question avec « *Œuvres anthumes* » d'Alphonse Allais, sous-titre d'un ouvrage de 1893.

b) Et en utilisant ci-dessus l'expression « *jouer sur les morphèmes* », sur quoi joue-t-on?

Exercice 3 Montrer que le suffixe *-eur* qui associe à un verbe X un nom désignant celui qui X-e (le marcheur est celui qui marche) n'est pas un syntaxème flexionnel.

Exercice 4 Quelle est, en français contemporain, la particularité des formes verbales *ci-git*, *j'ai ouï dire*, *oyez bonnes gens*?

Exercice 5 Soit le corpus de swahili suivant. Il s'agit de formes verbales qui forment à elles seules des phrases du swahili. En considérant que ces formes sont des combinaisons libres, déterminer les différent syntaxèmes qui les composent et en déduire un tableau de conjugaison du swahili. Pour extraire chacun des syntaxèmes, on commencera par repérer des paires minimales et les commutations associées.

```
1 atanipenda 'il m'aimera'
```

- 2 atampenda 'il l'aimera'
- 3 atawapenda 'il les aimera'
- 4 nitampenda 'je l'aimerai'
- 5 *utanipenda* 'tu m'aimeras'
- 6 tutampenda 'nous l'aimerons'
- 7 nitakupenda 'je t'aimerai'
- 8 *atanipiga* 'il me battra'
- 9 atampiga 'il le battra'
- 10 nitawapenda 'je les aimerai'
- 11 amenipiga 'il m'a battu'
- 12 amempiga 'il l'a battu'
- 13 *alikupiga* 'il te battait'
- 14 amekupiga 'il t'a battu'
- 15 atakupenda 'il t'aimera'
- 16 atatupenda 'il nous aimera'
- 17 *alimpiga* 'il le battait'
- 18 *utampenda* 'tu l'aimeras'
- 19 watampenda 'ils l'aimeront'
- 20 atakupiga 'il te battra'
- 21 ananipiga 'il me bat'
- 22 anampiga 'il le bat'
- 23 unamsumbua 'tu l'ennuies'
- 24 atakusumbua 'il t'ennuiera'
- 25 alinipiga 'il me battait'
- 26 anakupiga 'il te bat'

27 *tunakulipa* 'nous te payons' 28 *wametulipa* 'ils nous ont payés'

Exercice 6 Décomposer le plus long mot du français, anticonstitutionnel-lement, en morphèmes. Indiquer par des parenthésages dans quel ordre ces morphèmes peuvent se combiner. Étudier la diagrammaticité des différentes combinaisons.

Exercice 7 Montrer que le *-ible* de *lisible* est un allomorphe du *-able* de *adaptable*.

Exercice 8 Quel est le signifiant du morphème dérivationnel dans le mot *définition*? Comparer avec *compression* et *constitution*.

Exercice 9 Pourquoi peut-on parler d'allomorphie plutôt que de supplétion dans la relation entre les radicaux *exprim(er)* et *express(ion)*?

Exercice 10 Pourquoi peut-on considérer que le présent de l'indicatif est réalisé par un syntaxème zéro en français?

Exercice 11 On s'intéresse aux alternances de radical dans la conjugaison.

- a) Montrer que chacun des verbes aimer, rigoler et pleurer alternent deux formes orales du radical.
- b) Quelles sont les conditions qui contrôlent la distribution des deux formes des radicaux de ces verbes? Montrer que ces alternances sont entièrement déterminées par la phonologie du français.
- c) Comparer leur distribution avec celle des formes des verbes APPELER et MOURIR. Pourquoi s'agit-il d'allomorphie dans ce cas?

5.26 Syntaxème et morphologie



Lectures additionnelles

On peut faire remonter l'analyse morphologique aux tout premiers ouvrages en linguistique et les notions de radical et d'affixe sont déjà bien établies dans l'antiquité. L'intérêt pour la morphologie a été renouvelé au $20^{\rm e}$ siècle par l'étude de nouvelles langues et notamment les langues amérindiennes. Le livre d'Edward Sapir1921 est probablement l'ouvrage le plus marquant de cette époque.

La notion unificatrice de **morphème** est introduite dans **Bloomfield1933** au chaptire 10. La morphologie est traitée aux chapitres 12 et 13. On trouvera notamment la description des formes du masculin des adjectifs français par troncation des formes du féminin (chapitre 13). On pourra également consulter **Greenberg1954[1960]** qui propose une étude quantitative des constructions morphologiques dans 8 langues différentes et qui présente une définition du morphème et une classification des opérations morphologiques similaire à la notre.

La morphologie morphématique a été en partie rejetée au profit d'une morphologie lexématique (Beard1995, Fradin2003) (voir HaspelmathSims2013 pour une comparaison des deux approches). Si nous avons maintenu la notion de morphème dans notre présentation, nous le distinguons du syntaxème, dont nous faisons l'unité minimale de combinaison libre. La notion de syntaxème étend la notion de lexème en évitant d'introduire une frontière entre les unités du lexique et de la grammaire. Comme dans la morphologie lexématique, nous rejetons l'idée de morphème zéro, lui substituant la notion de syntaxème zéro, c'est-à-dire d'unité syntaxique qui n'a pas de réalisation phonologique et qui donc précisément ne correspond pas à un morphème. On pourra consulter l'ouvrage de Lemaréchal1997 consacré aux unités de ce type.

L'opération de **combinaison libre** \oplus est introduite par Igor Mel'čuk qui nomme cette opération l'*union linguistique*. Elle apparaît dans la plupart de ses ouvrages. Néanmoins, il n'en donne pas, à notre connaissance, de définition et la considère plutôt comme une primitive linguistique. Par ailleurs, il ne l'oppose pas, comme nous, à l'opération \boxplus de combinaison propre, elle

même opposée à l'opération + de combinaison impropre.

La notion de **liberté** telle que nous l'avons introduite peut être attribuée à **Martinet1985**, qui n'en donne pas de définition satisfaisante à notre avis, mais considère que « Les deux monèmes /dòn-/ et /-é/ de *donnait* sont des monèmes libres parce que *donnait* ne se comporte exactement, dans ses rapports avec le contexte, comme aucun monème unique de la langue. » (p. 34) Il n'y a pas néanmoins chez Martinet de distinction entre syntaxème et sémantème et son « synthème », dont il dit qu'il « se comporte vis-à-vis des autres monèmes de la chaîne comme un monème unique » (p. 37) (ce qui le rapprocherait du syntaxème), doit au final être rapproché de notre sémantème, puisqu'il se présente comme un « choix unique » (p. 36) (voir Chapitre 6 sur *Sémantèmes et syntaxèmes*).

Apothéloz2002 est une présentation très claire de la construction des lexèmes du français, illustrée par le lexique du français. L'émergence du morphème -eraie évoquée à la section ?? sur les morphèmes sous-lexicaux y est expliquée. On pourra également consulter Gardes-Tamine1990 pour une introduction à la morphologie. Pour une présentation monumentale de la morphologie, on consultera les 5 tomes du *Cours de morphologie générale* d'Igor Mel'čuk. Le volume 4 présente les morphèmes, l'alternance et la distinction entre mégamorphes forts et faibles. Enfin HaspelmathSims2013 est une présentation simple et assez complète de la morphologie.

Denis Apothéloz2002 La construction du lexique français, Ophrys.

Robert **Beard1995** *Lexeme-morpheme base morphology : a theory of inflection and word formation,* State University of New York Press.

Leonard **Bloomfield1933** *Language*, Holt, Rinehart and Winston, [traduction française : *Le langage*, Payot, 1970].

Bernard **Fradin2003** *Nouvelles approches en morphologie.* Presses universitaires de France.

Joëlle Gardes-**Tamine1990** *La grammaire, volume 1 : Phonologie, morphologie, lexicologie,* Armand Colin.

Joseph H. Greenberg (1954 [1960]). A quantitative approach to the morphological typology of language. In R. F. Spencer (ed.), *Method and Perspective in Anthropology: Papers in Honor of Wilson D. Wallis*, 192-220. Minneapolis: University of Minnesota Press. Reprinted in *International Journal of American Linguistics* 26(3),178-194.

Martin Haspelmath & Andrea D. Sims
2013 $Understanding\ morphology$, Routledge.

Alain Lemaréchal 1997, Zéro(s), Presses Universitaires de France.

André Martinet1985 Syntaxe générale, Armand Colin.

Igor Mel'čuk (1993-2000) *Cours de morphologie générale*, 5 volumes, Presses de l'Université de Montréal/CNRS Éditions.

Edward Sapir 1921 Language : An introduction to the study of speech, New York : Harcourt, Brace and cie, [traduction française de Solange-Marie Guillemen, Le language : introduction à l'étude de la parole. Payot, 1967].



Corrections des exercices

Corrigé 1 a) La décomposition est propre, car *national* est à *nation* ce que *régional* est à *région*.

b) *Idéal* n'est pas à *idée* ce que *national* est à *nation*. On ne voit pas très bien le lien sémantique entre *idée* et *idéal* et ce que pourrait être la contribution de *-al* dans *idéal*.

Corrigé 2 a) Le mot extime a été créé à partir d'intime et du couple intérieur-extérieur en appliquant le quatrième de proportionnel de Saussure. Une telle création joue sur les morphèmes en imposant une décomposition in+time et donc une analyse en deux morphèmes de intime que les locuteurs ne font normalement pas. Même chose avec anthume créé à partir de posthume et du couple antérieur-postérieur.

b) Nous jouons avec l'expression « jouer sur les mots ». Cette expression est une expression figée (voir section 6.8 sur *Phrasème*). La commutation

de *mots* par *morphèmes* est normalement illégitime et provoque un défigement de l'expression.

Corrigé 3 La combinatoire du suffixe -eur est irrégulière et imprévisible : certains verbes n'ont pas de dérivé en -eur (°aimeur, °arriveur), d'autres ont des dérivés avec des sens très spécifiques (rongeur désigne un type d'animal, descendeur s'utilise pour un skieur spécialiste de l'épreuve de descente, etc.). À l'inverse un lexème en -eur comme auteur n'est pas dérivé d'un verbe.

Corrigé 4 Les formes sont figées et il n'y a plus de combinaison libre du lexème verbal et de la flexion. Nous ne sommes plus capable de conjuguer ces verbes. Quel était l'infinitif de *git* ? (*gésir*!)

Corrigé 5 La décomposition en syntaxèmes repose sur l'hypothèse qu'une différence minimale de forme correspond à une différence minimale de sens ou de signification. Par exemple, pour 1 et 2, on suppose que la différence de forme (-ni- vs -m-) correspond à la différence de sens entre les énoncés, c'est-à-dire à un indice objet 1ère personne du singulier vs 3ème personne du singulier (il est préférable de parler ici d'indice pronominal plutôt que de pronom; voir la notion d'indice pronominal dans l'encadré?? Pronoms ou syntaxèmes flexionnels?). De même, pour 8 et 11, la différence de forme -ta- vs -me- correspond a priori au passage du pluriel au passé ou accompli. (La traduction en français est forcément approximative et ne permet de savoir quel est précisément le sens des syntaxèmes. En tout cas, il est clair que parler de passé composé pour 11 n'aurait aucun sens, car si la forme est composée en français, elle ne l'est évidemment pas en swahili.) En continuant à procéder ainsi, on vérifiera que la conjugaison du swahili est extrêmement régulière pour les formes considérées ici. Les formes verbales du swahili se composent ainsi de quatre syntaxèmes : indice sujet ⊕ temps \oplus indice objet \oplus radical verbal.

Corrigé 6 $anticonstitutionnellement = [anti \boxplus ([(con + stitu) \boxplus tion] \boxplus el)] \boxplus ment$. La décomposition de constitu en con + stitu est impropre, comme discutée en 2.2.4.

Corrigé 7 lisible est à lisons ce que adaptable est à adaptons. On en déduit directement que lisible = lis \boxplus ible, adaptable = adapt \boxplus able et -ible et -able sont des allomorphes.

Corrigé 8 Le radical du verbe DÉFINIR est *définiss-*/definis/, donc le suffixe de *définition* est -*ion*/jɔ/. Le suffixe est le même dans *compression*. Par contre, le suffixe de *constitution* est -*tion*/sjɔ/, puisque le radical du verbe CONSTITUER est *constitu-*.

Corrigé 9 Il existe tout une série de verbes en X-primer qui ont un nom dérivé en X-pression : comprimer, déprimer, exprimer, opprimer, réprimer, supprimer. Ceci permet de postuler que le quasi-morphème -prim- a un allomorphe -press-.

Corrigé 10 Une forme verbale comme *chantons* /ʃãtɔ/ s'oppose à *chantions* /ʃãtjɔ/ ou *chanterons* /ʃãtʁɔ/. Nous sommes dans un paradigme fermé de syntaxèmes et l'absence des morphèmes /j/ ou /ʁ/ est signifiante. Le présent de l'indicatif est donc un syntaxème sans réalisation phonologique, c'est-à-dire une syntaxème zéro.

Corrigé 11 Tous ces verbes ont une alternance vocalique : j'aime [zɛm] vs nous aimons [nuzemɔ] (voir aussi je cède [zəsɛd] vs nous cédons [nusedɔ]), je rigole [zəsɛd] vs nous rigolons [nusedɔ],

nasalization missing

je pleure [зәрlœв] vs nous pleurons [nupløвэ]

nasalization missing

, j'appelle [ʒapɛl] vs nous appelons [nuzapəlɔ]

nasalization missing

, je meurs [зәтœв] vs nous mourons [numuвэ]

nasalization missing

. Ces alternances ont la même distribution : un radical est utilisé devant un zéro et l'autre devant une voyelle. Pour les trois premiers verbes, il s'agit d'une alternance phonologiquement contrôlée entre $[\epsilon]$ vs [e], $[\mathfrak{o}]$

vs [o] et $[\infty]$ vs $[\emptyset]$ selon que le phonème (ou l'archiphonème) est dans une syllabe fermée (CVC) ou ouverte (CV) (voir l'encadré ?? Alternance vocalique en français : allomorphie ou alternance phonologique?). Pour les deux derniers verbes, l'alternance a certainement une origine phonologique, mais elle est aujourd'hui réalisée par deux phonèmes bien distincts ($[\varepsilon]$ vs $[\vartheta]$ et $[\infty]$ vs $[\upsilon]$) et il s'agit donc d'allomorphie.

6 Sémantèmes et syntaxèmes : Unités de sens vs unités de combinatoire

6.1 Arbitraire du sémantème

L'étude des sémantèmes est une question de **sémantique** qui dépasse les objectifs de ce livre. Mais, pour bien comprendre ce que sont les syntaxèmes, il nous semble utile de clarifier d'abord ce que sont les sémantèmes. Comme nous le verrons dans ce chapitre, l'extension des syntaxèmes, qu'elle soit prise dans sa dimension syntagmatique ou paradigmatique, est généralement comprise entre celle des morphèmes et celle des sémantèmes (voir notamment l'encadré ?? sur les *Extensions paradigmatique et syntagmatique*).

Notre définition des unités sémantiques est plus restrictive que celle des signes linguistiques. Revenons sur notre définition des signes donnée à la section 5.3 sur la Décomposition propre des signes. Nous avons vu que le signe broyeur se décompose proprement en broy \boxplus eur, car broyeur est à broyer ce que compresseur est à compresser. Les composantes broy- et -eur de broyeur peuvent donc se voir attribuer une contribution sémantique propre et sont ainsi des signes au sens plein du terme. Mais malgré cela, broyeur n'est pas compositionnel (voir définition dans la section suivante). Il y a quelque chose d'arbitraire dans le signe broyeur, tant au niveau du signifiant que du signifié. Au niveau du signifiant, pourquoi utiliset-on broyeur pour un appareil qui sert à broyer et pas laveur pour une machine qui sert à laver? Au niveau du signifié, on ne désigne pas par broyeur n'importe quel appareil qui sert à broyer : un hachoir à viande, qui broie plus qu'il ne hache, s'appellera toujours un hachoir.

Ce qui est vrai pour *broyeur* l'est aussi pour des combinaisons syntaxiques, comme *machine à laver*. La combinaison *machine à laver* peut paraître plus libre que *broyeur*, pourtant lorsqu'on fait des commutations sur *machine* ou *laver*, on obtient des combinaisons comme *appareil à laver* ou *machine à broyer* dont la nature est différente de celle de *machine à laver*. Une machine à laver n'est pas n'importe quelle machine qui sert à laver : *machine à laver* désigne un type particulier d'appareil ménager servant à laver le linge. On n'appellera pas *machine à laver* la machine qui sert à laver les voitures dans les stations services. Il y a

donc quelque chose d'**arbitraire** dans la relation entre le sens 'appareil ménager servant à laver le linge' et le signifiant *machine à laver*.

Les unités telles que *broyeur* ou *machine à laver* dont on peut trouver en partie le sens à partir du sens de leurs composants sont dites TRANSPARENTES. Les signes qui composent un signe complexe transparent ne sont pas pour autant nécessairement des unités sémantiques. Lorsqu'on parle de transparence, on raisonne dans le sens de l'analyse, du décodage des unités. Pour bien comprendre la spécificité des unités sémantiques parmi les signes linguistiques, il faut se placer dans le sens de la **synthèse**, de la production d'un énoncé à partir d'un sens.

6.2 Chaque sémantème suppose un choix

Nous proposons de caractériser les unités sémantiques à partir de la notion de CHOIX introduite par André Martinet. Voici ce qu'il en dit dans le chapitre intitulé *Chaque unité suppose un choix* de ses *Eléments de linguistique générale* (Martinet1960 : 26) :

« Soit un énoncé comme c'est une bonne bière /setynbonbier/. [...] Si nous sommes à même de dire quelque chose sur les latitudes combinatoires de /bɔn/, c'est que ce segment de l'énoncé a été reconnu comme représentant une unité particulière distincte de /yn/ et de /biɛr/. Pour arriver à ce résultat, il a fallu constater que /bɔn/, dans ce contexte, correspondait à un choix spécifique entre un certain nombre d'épithètes possibles; la comparaison d'autres énoncés français a montré que dans les contextes où figure /bɔn/ on trouve aussi /eksɛlãt/ (excellente), /movɛz/ (mauvaise), etc. Ceci indique que le locuteur a, plus ou moins consciemment, écarté tous les compétiteurs qui auraient pu figurer entre /yn/ et /biɛr/, mais qui ne se trouvaient pas convenir en l'occurrence. Dire de l'auditeur qu'il comprend le français implique qu'il identifie par expérience les choix successifs qu'a dû faire le locuteur, qu'il reconnaît /bɔn/ comme un choix distinct de /yn/ et de celui de /biɛr/, et qu'il n'est pas exclu que le choix de /bɔn/ au lieu de /movɛz/ influence son comportement. » (Dans cette citation, nous avons modifié les transcriptions phonémiques et utilisé les conventions de cet ouvrage.)

Prenons l'exemple de l'énoncé *Pierre mange une pomme de terre*. Dans cet énoncé, les mots *pomme*, *de* et *terre* sont des signes linguistiques, mais aucun d'eux ne résulte d'un choix.

Par CHOIX, nous entendons les choix lexicaux et grammaticaux que fait le locuteur lorsqu'il cherche à **exprimer un sens** qu'il veut communiquer.

Lorsque le locuteur produit *pomme de terre*, *pomme* n'a pas été choisi par opposition à *poire* ou *banane*, *de* n'a pas été choisi par opposition à *à* ou *dans*, et *terre* n'a pas été choisi par opposition à *eau* ou *feu*. C'est bien *pomme de terre* dans son intégralité qui a été choisi par opposition à *carotte*, *chou-fleur* ou *haricot vert*. Ainsi *pomme de terre*, *carotte*, *chou-fleur* et *haricot vert* forment un paradigme de choix ou système d'opposition et chacun de ces choix est indivisible. Il en résulte que *pomme de terre* est ici une unité sémantique et les segments qui la composent n'en sont pas (dans cet énoncé).

Nous appelons unité sémantique tout signe qui résulte d'un ou plusieurs choix du locuteur.

Il existe des signes comme *machine* dans *machine* à *laver* qui ne forment pas une unité sémantique, mais seulement une composante d'une unité sémantique. Il existe aussi des signes comme les syntaxèmes d'accord en genre des adjectifs avec le nom ou le syntaxème d'infinitif qui ont une signification purement grammaticale et ne sont donc ni des unités sémantiques, ni même des composantes d'une unité sémantique. De tels signes ne résultent pas de choix du locuteur; ils sont entièrement requis par la grammaire de la langue. Nous développons ce point dans la section 6.4 sur le *Syntaxème vide*.

Nous appelons sémantème tout signe qui résulte d'un choix indivisible. Les sémantèmes sont les unités sémantiques minimales : ils ne peuvent pas être décomposés en deux unités sémantiques.

Ainsi *pomme de terre* est une unité sémantique, mais ses composantes *pomme*, *de* et *terre* n'en sont pas. Il en résulte que *pomme de terre* est un sémantème. Il en va de même pour *broyeur* ou *machine à laver*.

Une unité sémantique qui peut être décomposée en deux unités sémantiques est dite COMPOSITIONNELLE.

Par exemple, une unité comme *livre de syntaxe* est compositionnelle. Ici *livre* et *syntaxe* sont choisis librement et peuvent commuter avec des quasi-synonymes comme *bouquin* ou *grammaire* (*un bouquin de syntaxe*, *un livre de* grammaire) sans que la variation de sens soit plus importante qe la variation de sens qui existe ailleurs entre *livre* et *bouquin* ou entre *syntaxe* et *grammaire* (*j'ai acheté un livre/bouquin, j'aime la syntaxe/grammaire*).

6.3 Motivation du signe

Un sémantème qui est composé de plusieurs signes est dit COMPLEXE (voir la section 2.2.14 Signème libérable, radical, affixe). L'existence de sémantèmes complexes, comme broyeur ou machine à laver, complexifie le système linguistique, puisqu'il n'y a plus correspondance systématique entre les segments minimaux du point de vue du sens (les sémantèmes) et les segments minimaux du point de vue de la forme (les morphèmes). Mais d'un autre côté, l'utilisation de sémantèmes complexes offre une certaine économie en réduisant le caractère arbitraire du lien entre la forme et le sens. Cette relative motivation du signifiant des sémantèmes complexes a été bien dégagée par Saussure1916 :

« Le principe fondamental de l'arbitraire du signe n'empêche pas de distinguer dans chaque langue ce qui est radicalement arbitraire, c'est-à-dire immotivé, de ce qui ne l'est que relativement. Une partie seulement des signes est absolument arbitraire; chez d'autres intervient un phénomène qui permet de reconnaître des degrés dans l'arbitraire sans le supprimer : *le signe peut être relativement motivé*.

Ainsi *vingt* est immotivé, mais *dix-neuf* ne l'est pas au même degré, parce qu'il évoque des termes dont il se compose et d'autres qui lui sont associés, par exemple *dix*, *neuf*, *vingt-neuf*, *dix-huit*, *soixante-dix*, etc.; pris séparément *dix* et *neuf* sont sur le même pied que *vingt*, mais *dix-neuf* présente un cas de motivation relative. Il en est de même de *poirier*, qui rappelle le mot simple *poire* et dont le suffixe *-ier* fait penser à *cerisier*, *pommier*, etc.; pour *frêne*, *chêne*, etc., rien de semblable. »

Le fait que les sémantèmes complexes soient en partie motivés n'enlève pas totalement le caractère arbitraire de la combinaison de morphèmes qui est retenue parmi toutes celles possibles. Ainsi pour *machine à laver*, les Québécois disent *laveuse* et les Français disent aussi *lave-linge*, mais à côté de *aspirateur*, nous n'avons ni *machine à aspirer*, ni *aspire-poussière*. On peut aussi noter que les sémantèmes choisis pour désigner un même objet peuvent exprimer des points de vue différents sur cet objet, comme l'illustre les « synonymes » *téléski*, *remonte-pente* et *tire-fesse* : le même appareil est présenté successivement comme un appareil qui permet de se déplacer avec ses skis, comme un appareil qui permet de se déplacer par une traction au niveau des fesses. À l'inverse, des quasi « paraphrases » comme *ventilateur* et *moulin à vent* désignent des objets très différents. C'est ce côté arbitraire du choix de la combinaison retenue parmi tous les possibles qui manifeste le caractère figé des sémantèmes.

La motivation ou transparence doit être distinguée de la diagrammaticité (voir la section 5.3 sur la *Décomposition propre des signes*). Il s'agit en quelque sorte du point de vue inverse.

Une combinaison AB est transparente si les sens de A et B permettent d'avoir une idée du sens de AB.

A l'inverse, la combinaison AB est diagrammatique si A et B commutent suffisamment proprement dans la combinaison AB pour qu'on puisse attribuer des sens à A et B.

Bien qu'indépendantes, les deux notions sont liées. La diagrammaticité de BROYEUR et COMPRESSEUR permet d'attribuer un sens au suffixe -eur. Et comme ce suffixe a acquis un sens, les sémantèmes complexes BROY+EUR et COMPRESS+EUR en deviennent transparents. On comprend mieux la différence entre ces notions quand on considère un suffixe rare comme celui de ROYAUME. Ce sémantème est très diagrammatique, car il entre dans un paradigme avec principauté ou duché et on peut donc voir qu'il est la combinaison de roi et -aume et attribuer un sens au suffixe -aume, mais il n'est pas très transparent, car le suffixe -aume n'apparaît pas ailleurs et son sens ne préexiste pas à la connaissance du sens de rovaume. À l'inverse, une locution comme LEVER LE COUDE 'boire de l'alcool' est assez motivée (et donc en partie transparente), puisque pour boire il faut lever le coude, mais elle n'est pas du tout diagrammatique, car la connaissance du sens de la locution ne permet pas d'attribuer des parties de ce sens aux composantes de la locution. On peut encore noter que les locuteurs tendent à produire de la diagrammaticité et à projeter des parties du sens d'une locution sur les syntaxèmes qui la composent. C'est pas exemple ce qu'on observe avec une locution comme PRENDRE LE TAUREAU PAR LES CORNES 'commencer à résoudre un problème', où TAUREAU peut s'attribuer le sens 'problème' comme dans l'exemple suivant : La ministre de l'Emploi, Monica De Coninck, a manifestement décidé de prendre le taureau de l'emploi des quinquas par les cornes (lesoir.be). Cette diagrammatisation peut aller jusqu'à rendre l'expression compositionnelle, comme poser un lapin, où lapin est aujourd'hui devenu un sémantème indépendant (C'est le deuxième lapin de la semaine!).

6.4 Syntaxème vide

Un syntaxème vide est un syntaxème sans contribution sémantique, qui n'a pas fait l'objet d'un choix, mais a été imposé par la grammaire de la langue. Un tel syntaxème ne fait donc pas partie d'un sémantème.

Les grammèmes d'accord en genre présentent un cas intéressant de syntaxèmes vides. Dans le syntagme *la petite fille* /lapətitfij/, l'article et l'adjectif s'accordent en genre avec le nom fille. Ces accords, /-a/ pour l'article et /-ə/ pour l'adjectif (c'est-à-dire l'absence de troncation de la consonne finale de l'adjectif, voir l'encadré ?? sur *Syntaxème zéro et troncation en français*), sont des syntaxèmes flexionnels, que nous appelons des GRAMMÈMES D'ACCORD. Ces grammèmes ne sont pas des sémantèmes : ils n'expriment pas un choix et ont uniquement un rôle syntaxique de marquage de la dépendance entre deux mots. En français, il existe deux syntaxèmes flexionnels de genre, le féminin et le masculin. Ils se combinent avec les grammèmes de nombre et sont spécifiques aux adjoints du nom, c'est-à-dire les adjectifs et les déterminants. Le signifié des grammèmes d'accord est une signification grammaticale qui n'est pas porteuse de sens. Les grammèmes d'accord sont des syntaxèmes vides. (Voir néanmoins le cas d'accord entre grammèmes dans l'encadré qui suit.)

Rappelons au passage que le genre est un *trait catégoriel* du nom qui contrôle sa combinatoire et la nature des syntaxèmes d'accord en genre sur certains éléments avec lesquels il se combine (les adjectifs, les déterminants et les participes) (voir le Chapitre ?? sur les *Catégories nanosyntaxiques*). Certains noms possèdent une alternance de « sexe » réalisée par un affixe dérivationnel, qui n'est pas un syntaxème (voir l'encadré ?? sur les signes *À la limite entre morphème sous-lexical et syntaxème flexionnel*).

La notion de signe vide terme a été introduite par Lucien Tesnière qui parle de *mot vide* (**Tesnières1959** : Chapitre 28 — Mots pleins et mots vides) :

- « 1. Il y a deux espèces de mots essentiels, les mots pleins et les mots vides.
- 2. Les mots pleins sont ceux qui sont chargés d'une fonction sémantique, c'est-à-dire ceux dont la forme est associée directement à une idée, qu'elle a pour fonction de représenter et d'évoquer. [...]
- 3. Les mots vides sont ceux qui ne sont pas chargés d'une fonction sémantique. Ce sont de simples outils grammaticaux ([Note] Damourette et Pichon disent très heureusement des « struments », d'autres ont proposé « mots-charnières ».) dont le rôle est uniquement d'indiquer, de préciser ou de transformer la catégorie des mots pleins et de régler leurs rapports entre eux. »

Parmi les syntaxèmes vides, outre les grammèmes d'accord, on considère :

les explétifs comme le sujet IL de *il pleut*, qui n'a pas de contribution sémantique et sert uniquement à remplir la contrainte syntaxique de présence d'un sujet;

- les translatifs purs (voir la section ?? sur La translation syntaxique) comme la conjonction de subordination QUE dans je sais que tu viens : ce syntaxème sert uniquement à permettre à un verbe d'occuper une position syntaxique où un substantif est attendu; les pronoms relatifs, comme le que de le livre que je lis, sont aussi d'une certaine façon des syntaxèmes vides, comme on le verra dans le Chapitre ?? sur l'extraction;
- les syntaxèmes marquant les **régimes**, comme le À de *je parle à quelqu'un*, sont aussi généralement considérés comme des syntaxèmes vides, même si ce cas est plus discutable (voir l'encadré qui suit).

6.5 Constructions verbales et accords : signes vides?



Dans Pierre parle à Marie, nous considérons que la préposition à n'est pas une unité sémantique, mais qu'elle fait partie du régime imposé par PARLER. Autrement dit, la préposition à ne constitue pas un choix séparé de celui de PARLER: plus exactement, le choix fait ici est celui d'une acception particulière de PARLER qui se construit avec un complément d'objet indirect introduit par à, le verbe et sa construction formant un tout indissociable.

La version extrême de cette position est de considérer que à est un signe vide, c'est-à-dire sans contribution sémantique (voir la section précédente). La position inverse est au contraire de considérer que le Parler de Pierre parle à Marie est le même que le PARLER de Pierre parle en dormant ou Pierre parle anglais et que la nuance de sens qu'il y a entre ces trois occurrences de PARLER est due à la construction avec laquelle PARLER se combine. Autrement dit, lorsque Pierre parle à Marie, le sens sous-jacent qui est que 'Pierre s'adresse à Marie' vient davantage de la construction « X V à Y », dans laquelle s'adresser, parler, téléphoner ou ÉCRIRE peuvent occuper la position V, que du verbe qui occupe cette position. Bien sûr cette construction signifiant en elle-même 's'adresser à quelqu'un' n'est compatible qu'avec des verbes du type s'Adresser et il est donc difficile de savoir d'où vient exactement la contribution sémantique. En admettant que le lexème PARLER se combine avec une construction comme il se combine avec sa flexion, il reste que la combinaison d'un lexème et d'une construction est une combinaison liée : par exemple, si on reprend les trois verbes écrire, parler et téléphoner qui se combinent tous avec « X V à Y », on voit qu'on peut dire parler anglais, mais pas *téléphoner anglais, ni *écrire anglais (on dit écrire en anglais). On peut dire écrire quelque chose à quelqu'un, mais pas parler quelque chose à quelqu'un. Par contre, on peut parler de quelque chose, mais pas *écrire de quelque chose ou *téléphoner de quelque chose. Donc les V qui se combinent avec une construction



donnée ne commutent pas librement avec d'autres V.

En conclusion, il peut être envisageable de considérer la construction du verbe comme un signe à part entière. On peut même considérer que cette construction est elle-même la combinaison de plusieurs signes, chacun correspondant à la relation entre le verbe et un de ses actants (voir le Chapitre ?? sur les Fonctions syntaxiques). La préposition ne serait donc pas ici un mot vide, mais le signifiant d'un signe indiquant une relation de destinataire entre l'action décrite par le verbe et son dépendant : dans *Pierre parle à Marie*, Pierre parle — il profère des paroles — et Marie en est le destinataire. Néanmoins la combinaison entre le verbe et sa construction est une combinaison liée : le paradigme des lexèmes appelant une construction donnée n'est pas régulier et, inversement, pour un lexème donné, les constructions dans lesquels il entre est restreint et peu prévisible. (Voir la section 6.11 Collocation et choix liés pour les combinaisons liées de deux sémantèmes.)

Nous avons vu que les syntaxèmes d'accord en genre sont des syntaxèmes vides. Un exemple comme Les animaux boivent /lezanimobwav/, opposé à L'animal boit /lanimalbwa/, est plus problématique. Il y a bien ici une unité sémantique de pluriel, mais il y a trois syntaxèmes exprimant le pluriel /-ez/, /-o/ et /-ə/, associés respectivement à l'article, au nom et au verbe. Un linguiste comme Martinet considère qu'il s'agit d'un sémantème possédant un signifiant discontinu, tandis qu'un linguiste comme Mel'čuk, considère 1) que l'un des trois syntaxèmes est un sémantème et que les deux autres sont des syntaxèmes d'accord, et 2) que c'est le nom qui se combine avec le sémantème. Cette dernière proposition est motivée par le fait que le sens du sémantème de pluriel porte effectivement sur le nom et permet de décider si le nom désigne un seul référent ou plusieurs. Par ailleurs, il existe des noms qui requièrent essentiellement le singulier, comme EAU, CHOCO-LAT ou PEUR, et d'autres toujours le pluriel, comme RILLETTES ou FIANÇAILLES, et on voit dans de tels cas que le nom forme un phrasème (voir définition à la section 6.8 qui suit) avec le nombre et que c'est lui qui va déclencher l'accord des autres éléments lexicaux. Mais l'on pourrait rétorquer que la plupart des déterminants portent intrinsèquement un nombre. Une troisième proposition serait donc de considérer que c'est le déterminant qui porte réellement le nombre soit dans son sens lexical, comme pour DEUX, soit sous la forme d'un grammème de nombre qui se combine avec le déterminant proprement dit, comme pour l'article. Cette option est renforcée par le fait que le nombre n'est généralement marqué à l'oral que sur l'article (Les vaches broutent /levasbrut/, opposé à La vache broute /lava[brut/). Pour conclure, nous dirons que les trois analyses s'accordent de toute façon sur le fait qu'il y a un sémantème et plusieurs syntaxèmes. Reste

la question de savoir si ce sémantème correspond à l'un de ces syntaxèmes ou à leur combinaison, question dont il n'est pas sure qu'elle soit pertinente.

6.6 Décomposition en sémantèmes

La première étape dans l'analyse sémantique d'un énoncé est d'en repérer les sémantèmes. Considérons l'énoncé :

(1) La moutarde me monte au nez.

La production de cet énoncé, dont le sens est à peu près 'je sens la colère monter en moi', suppose quatre choix. Cet énoncé contient donc quatre sémantèmes. Nous incitons le lecteur à arrêter un instant sa lecture et à chercher quelles sont ces quatre unités sémantiques.

La première est me, choisie par opposition à te, lui, etc. Plus exactement, il s'agit du lexème мої, la forme clitique me étant imposée par le prédicat verbal qui réalise son premier actant comme objet indirect. Dans une paraphrase comme Je commence à être en colère, moi sera réalisé par la forme sujet je. La deuxième unité sémantique est la locution la moutarde mont- au nez, que nous notons [⊠]LA MOUTARDE MONTER AU NEZ^N, signifiant 'sentir la colère monter en soi'. Comme dans le cas de pomme de terre, aucun des syntaxèmes qui composent cette locution ne peut être commuté individuellement : #Le wasabi me monte au nez, #La moutarde me grimpe au nez, #La moutarde me monte aux yeux. Aucune de ces phrases n'évoque la colère ou alors par un jeu de mots qui permet au locuteur de faire le lien avec La moutarde me monte au nez. Le choix de la moutarde montau nez est un choix unique et indivisible. La troisième unité sémantique est le présent, réalisé par -e (un syntaxème zéro à l'oral) et choisi par opposition à l'imparfait ou au futur (La moutarde me montait au nez, La moutarde me montera au nez). La quatrième unité sémantique est non segmentale : c'est la déclarativité. Notre énoncé s'oppose ainsi à La moutarde me monte-t-elle au nez? La déclarativité s'exprime ici par l'utilisation d'une proposition avec un verbe à l'indicatif et une courbe intonative descendante en fin de phrase, typique des énoncés déclaratifs (transcrit à l'écrit par l'utilisation du point final « . » par opposition à «? » ou «! »).

Nous étudierons la façon dont les sémantèmes se combinent dans le Chapitre ?? sur la *Structure syntaxique profonde*. On peut aussi consulter le Chapitre 2 où nous avons montré comment *Produire un énoncé* à partir d'une représentation sémantique.

Q

6.7 Pronoms **Q**

Les pronoms sont un autre type d'unités que l'on tend à opposer aux lexèmes pleins (voir la section 6.4 sur le *Syntaxème vide*). La plupart des pronoms sont bien des unités ayant une réelle contribution sémantique, mais ils contrastent effectivement avec des sémantèmes comme Arbre, voir, rouge ou dans. La plus grande partie des sémantèmes ont un sens dénotationnel : ils dénotent certaines entités ou situations du monde, c'est-à-dire que leur sens sélectionne intrinsèquement une partie des entités ou situations envisageables dans le ou les mondes imaginaires que nous construisons avec nos discours. On peut ainsi parmi les entités du monde identifier un sous-ensemble d'entités que l'on peut dénommer des *arbres* ou que l'on peut considérer comme *rouge*. De la même façon, parmi les situations mettant en jeu deux entités X et Y, on peut distinguer celles où X *voit* Y ou celles ou X est *dans* Y et cette décision est indépendante de la situation d'énonciation, c'est-à-dire de qui parle à qui, où et quand.

Cette « dénotationalité » contraste avec la propriété des pronoms, comme CELUI-CI, L'AUTRE, CA, MOI ou ICI. Les pronoms sont des unités qui permettent également de sélectionner une partie des entités ou des situations envisageables, mais cette sélection dépend complètement de la situation d'énonciation. Il n'y a pas d'entité du monde dont on puisse dire qu'elle est intrinsèquement *l'autre*, *ça*, *ici* ou *moi* : 'moi' est celui qui énonce (*Regarde-moi*!), 'ici' est le lieu où a lieu l'énonciation (*Ici*, *il fait plus chaud qu'à Paris.*), 'ça' est quelque chose que l'on montre (*Prends ça*!) ou dont on a parlé (*Ça m'a plu.*) et 'l'autre' se définit par rapport à un 'celui-là' déjà identifié (*Je vais plutôt prendre l'autre*.).

On distingue deux fonctions pour les pronoms. Un pronom est dit déictique si la ou les entités auxquelles il réfère peuvent être déterminées par la situation d'énonciation, indépendamment du discours qui entoure l'énonciation du pronom : Moi, toi, ici ou demain sont typiquement des déictiques ou celui-ci quand il s'agit d'une chose que l'on désigne de la main. Un pronom est dit anaphorique si la ou les entités auxquelles il réfère sont déterminées par le renvoi à une portion antérieure du discours, que l'on appelle l'antécédent du pronom. Dans les exemples suivants, le pronom anaphorique est en gras et son antécédent souligné :

- (2) Pierre a pris un livre. Il était sur la table.
- (3) Pierre a pris le livre qui était sur la table.
- (4) Pierre est allé en vacances à la montagne. Ca lui a plu.

On ajoute encore dans la catégorie des pronoms d'autres éléments qui n'ont pas de sens dénotationnel : les pronoms interrogatifs comme qui (**Qui** a parlé à Marie? Pierre, je crois.) ou les pronoms négatifs comme rien (Il n'a rien vu).

Les fonctions anaphorique et déictique ne sont pas la spécificité des pronoms :

- (5) Prends ce livre!
- (6) Pierre a pris un livre. Ce livre était sur la table.
- (7) Pierre a fait le tour du monde. Le voyage lui a plu.

Les pronoms partagent beaucoup de propriétés avec les déterminants et un même lexème possède souvent une acception qui est déterminant (*Plusieurs personnes sont venues*) et une autre qui est pronom (*Plusieurs sont venus*). La particularité des pronoms est de pouvoir référer sans s'appuyer sur un nom, contrairement aux déterminants qui doivent se combiner avec un nom. (Voir aussi la possibilité de traiter le pronom et le déterminant comme un même sémantème à la section ?? *Déterminant comme tête*?).

Pour les éléments qui ont spécifiquement une valeur anaphorique (ou déictique) mais qui ne reprennent pas un nom, on parle de PROFORME. Les verbes peuvent avoir des proformes, mais cela reste marginal dans la diversité des langues ; citons le verbe modal do pour l'anglais :

- (8) Do you plan to come with us? 'Prévois-tu de venir avec nous?'
- (9) Yes, I do. 'Oui, je le prévois', litt. 'Oui, je fais'.

6.8 Phrasème

Comme nous l'avons vu pour *broyeur* ou *machine à laver*, il est assez courant qu'un sémantème soit réalisé par une combinaison de morphèmes. Il s'agit même de la majorité des sémantèmes.

Un sémantème composé de plusieurs morphèmes est appelé un sémantème complexe. Un sémantème commute généralement librement avec son contexte (voir le cas particulier des collocations dans la section 6.11 *Collocation et choix lié*). Il s'agit donc soit d'un syntaxème, soit d'une combinaison de syntaxèmes.

Un sémantème qui est la combinaison de plusieurs syntaxèmes est appelé un phrasème. Les phrasèmes sont encore appelés des locutions ou des expressions figées dans la tradition grammaticale française et des expressions idiomatiques (angl. idioms) dans la tradition anglo-saxonne. Les délimiteurs $^{\boxtimes}$... $^{\boxtimes}$ servent à indiquer qu'une combinaison de syntaxèmes forme un phrasème.

On distingue parmi les sémantèmes complexes ceux qui sont des syntaxèmes, comme les lexèmes complexes blessure, aspirateur ou construction, et ceux qui sont des combinaisons de syntaxèmes, comme les phrasèmes $^{\boxtimes}$ se retourner dans sa tombe $^{\boxtimes}$ ou $^{\boxtimes}$ tiré par les cheveux $^{\boxtimes}$. Les signes qui composent ces deux types de sémantèmes complexes sont de nature différente. Les signes lexicaux qui apparaissent dans un lexème complexe, comme *bless*- dans blessure ou *aspir*- dans aspirateur, sont des lexèmes qui ont perdu leur combinatoire (*bless*- et *aspir*- ne se comportent plus comme les verbes bless(er) et aspir(er) dans ce contexte), mais ont en grande partie conservé leur sens : ce sont des lexèmes désyntactisés. À l'inverse, les signes lexicaux qui apparaissent dans les phrasèmes conservent leur combinatoire syntaxique, mais sont dépossédés de leur contribution sémantique usuelle : ce sont des lexèmes désémantisés.

Les phrasèmes sont par définition des sémantèmes et font donc l'objet d'un choix indivisible. Il n'y a donc pas de paradigme de choix sur les syntaxèmes qui les composent. C'est ce qu'on nomme le figement ou la lexicalisation de la combinaison. Une expression figée a **perdu sa compositionalité**.

Certains phrasèmes sont polymorphes : c'est bien sûr le cas lorsque les morphèmes sont eux-mêmes polymorphes comme $^{\boxtimes}$ s'en aller $^{\boxtimes}$, où le réfléchi (noté se) comme aller ont de nombreux allomorphes : je m'en vais, nous nous en irons, etc. Mais il y a aussi des phrasèmes qui ne s'expriment pas de manière univoque par une seule combinaison de lexèmes : c'est le cas par exemple de $^{\boxtimes}$ CHIER/FAIRE DANS SON FROC/PANTALON $^{\boxtimes}$ 'manifester une peur intense' avec deux alternances possibles et donc quatre formes en tout.

6.9 Lexie et grammie

De même que parmi les syntaxèmes nous avons distingué les lexèmes et les grammèmes, nous distinguons parmi les sémantèmes les lexies et les grammies.

Un sémantème qui commute avec des lexèmes, comme $^{\boxtimes}$ se retourner dans sa tombe $^{\boxtimes}$, est appelé une unité lexicale ou lexie.

Nous distinguons très nettement les termes *lexème* et *lexie*. En général, un lexème possède plusieurs acceptions qui constituent autant de lexies (voir la section 6.14 sur la *Délimitation paradigmatique des sémantèmes*). Un lexème peut aussi être partie d'un phrasème et ce sont alors les acceptions de ce phrasème qui constituent des lexies. La lexie est ainsi l'unité de nature lexicale minimale du point de vue de son signifié, tandis que le lexème est l'unité de nature lexicale minimale du point de vue de son syntactique.

Nous appellerons unité grammaticale ou grammie tout sémantème qui commute avec des syntaxèmes flexionnels. Il peut s'agir d'acceptions de syntaxèmes flexionnels, mais aussi de combinaisons de grammèmes comme le conditionnel $^{\boxtimes}$ -r- t-i- $^{\boxtimes}$ (dans *Nous aimerions vous rencontrer*) ou de combinaison d'un grammème et d'un lexème comme le passé composé $^{\boxtimes}$ AVOIR + participe_passé $^{\boxtimes}$.

Certaines lexies ont plutôt un rôle structurel et ne commute pas vraiment avec des lexèmes ou des grammèmes, comme le « cliveur » $^{\boxtimes}$ c'est ... Qu- $^{\boxtimes}$ (*C'est Pierre qui viendra*, *C'est* à Pierre que je l'ai dit) ou le présentatif $^{\boxtimes}$ Il y a ... Qui $^{\boxtimes}$ (*Il y a quelqu'un qui a sonné*, *Il y a ma voiture qui est en panne*). Nous les appelons des constructions de Phrase. Nous les distinguons des constructions verbales, qui représentent le régime d'un verbe et ses redistributions.

6.10 Verbes supports et unités grammaticales 🍳

Les grammies marquant des rôles syntaxiques sont souvent en concurrences avec des CONSTRUCTIONS À VERBES SUPPORTS. On appelle VERBE SUPPORT un verbe qui n'a pas de contribution sémantique propre et joue essentiellement un rôle syntaxique en permettant à un nom prédicatif d'occuper une position verbale. (Les verbes supports sont appelés *light verbs* 'verbes légers' dans la littérature anglo-saxonne.)

Par exemple, pour exprimer le sens 'gifler', on a le choix entre le verbe GIFLER et une construction à verbe support avec le nom GIFLE. Dans les deux cas, trois diathèses sont possibles :

(10) a. actif
Marie gifle Pierre Marie **donne** une gifle à Pierre

- b. passif avec agent
 Marie est giflée par Pierre Marie reçoit une gifle de Pierre
- c. passif sans agent Marie **se fait** gifl**er** Marie **se prend** une gifle

Le verbe support n'est pas sémantiquement vide : dans l'action de gifler, il y a bien un donneur et un receveur et les verbes donner, recevoir ou prendre sont sémantiquement motivés et ce sont des sémantèmes. Néanmoins la contribution des verbes supports consiste essentiellement à attribuer les rôles sémantiques aux différents participants comme le font les voix (voir définition dans le Chapitre ??): l'actif, le passif (être + participe_passé) et le passif sans agent (se faire + infinitif). C'est ce qui nous autorise à dire que les verbes supports n'ont pas de contribution sémantique propre.

Les verbes supports sont un cas particulier de **translatif** (voir la section ?? sur *La translation syntaxique*) et un cas particulier de **collocatif** (voir la section 6.11 qui suit).

6.11 Collocation et choix liés

Si l'unité sémantique AB est une **combinaison libre** $A \oplus B$, cela signifie que A et B sont nécessairement des unités sémantiques choisies indépendamment l'une de l'autre, dans des paradigmes de choix indépendants, et donc que l'unité AB est compositionnelle. Si l'unité sémantique AB est une **combinaison liée** A + B, il est possible que A et B soient tout de même des unités sémantiques, c'est-à-dire que A et B résultent de **choix distincts**. Néanmoins, dans ce cas, comme nous allons le voir, les choix de A et B ne peuvent pas être totalement indépendants.

Une collocation ou semi-phrasème est une combinaison non libre de deux unités sémantiques.

Dans l'énoncé *Pierre a peur*, le choix de PEUR s'oppose à ceux de FAIM, FROID ou ENVIE, tandis que le choix de AVOIR s'oppose à ceux de PRENDRE ou FAIRE. Le principe de commutation s'applique et la commutation est propre. On a bien deux unités sémantiques. Par contre, PEUR ne peut pas commuter dans cet énoncé

avec ses synonymes crainte ou effroi et à la place de *avoir crainte ou de *avoir effroi, on dira plutôt craindre ou être effrayé. De même, le verbe avoir ne peut pas commuter ici avec un autre verbe que prendre ou faire. La combinaison n'est donc pas libre.

Dans une collocation, les deux choix ne peuvent être libres et indépendants, sinon la combinaison serait nécessairement libre. Il s'agit de choix liés. C'est pourquoi une collocation est dite SEMI-FIGÉE.

Plus exactement, une collocation est toujours asymétrique : l'un des choix est dépendant de l'autre. Par exemple, dans Avoir peur, c'est peur qui est choisi librement, tandis qu'Avoir est choisi pour permettre a peur de jouer un rôle verbal (c'est un verbe support; voir l'encadré ?? qui précède) et ce choix est dépendant de peur, dans le sens où un autre nom de sentiment entraînerait un choix différent.

Le sémantème qui est choisi librement est appelé la BASE de la collocation. Le sémantème dont le choix dépend de la base est appelé un COLLOCATIF de cette base. On reconnaît un collocatif au fait qu'il ne peut pas se combiner avec certains sémantèmes qui peuvent commuter avec sa base dans d'autres contextes et qui sont pourtant sémantiquement proches.

Le tableau suivant met en parallèle les choix contraints par PEUR et COLÈRE pour l'expression d'un certain nombre de sens très généraux comme 'causer', 'commencer' ou 'intense'. (Une case vide indique l'absence de collocatif pour exprimer ce sens avec cette lexie.)

	PEUR	COLÈRE
'éprouver'	avoir peur	être en colère
'commencer à éprouver'	prendre peur	se mettre en colère
'causer'	faire peur à qqn	mettre qqn en colère pro- voquer la colère de qqn
'manifester (par un symptôme)'	trembler de peur	
[⊠] faire dans son froc [⊠] être	suffoquer de colère	
paralysé (par la peur)	bouillir (de colère)	
'éprouver intensément'	avoir une peur bleue être	être dans une colère noire
	vert/mort de peur	être rouge de colère
'éprouver faiblement'	avoir une petite peur	_
'décharger'	se libérer de sa peur	passer sa colère sur qqn
'qui ne se manifeste pas'	1	colère sourde, rentrée
'qui n'est pas contrôlé'	peur panique	colère aveugle

On peut constater que les choix possibles sont très différents. On aura noté en particulier le choix totalement arbitraire des adjectifs de couleur, *bleu* et *vert* pour la peur, *rouge* et *noir* pour la colère. Et si on peut avoir *une peur bleue*, on ne peut être *bleu de peur. On doit être *vert de peur*, mais on n'aura pas *une peur *verte*. Le fait qu'aucun de ces sémantèmes ne peut se combiner à la fois avec PEUR et COLÈRE montre bien qu'il s'agit de collocatifs.

Si la collocation partage avec le phrasème le caractère arbitraire de son signifiant, une collocation A + B se distingue d'un phrasème par la possibilité de commutation sur le collocatif (comme on l'a vu ci-dessus pour *avoir peur*) et par la possibilité de modifier indépendamment A et B (cf. *ne pas avoir peur* vs *avoir une belle peur*). La transparence sémantique ne permet pas par contre de les distinguer. Comparons par exemple les expressions *noyer le poisson* et *avoir les boules*: à première vue, aucune des deux expressions n'est très transparente, si l'on s'en tient au sens usuel des différents lexèmes qu'elles contiennent. Pourtant, alors que la première ne permet aucune commutation propre, la deuxième permet des commutations propres sur Avoir: *foutre les boules* à *qqn*, *C'est les boules*!, *Putain, les boules*!. Comme par ailleurs, *les boules* ne permet pas de commutation de *les* dans ce contexte (*#avoir la boule*, *#avoir des boules*), on en déduit que LES BOULES est un sémantème et que Avoir est un collocatif (un verbe support pour être plus précis). Au final, *avoir les boules* apparaît comme une semi-phrasème, tandis que NOYER LE POISSON est vraiment un phrasème.

Nous avons évoqué, dans l'encadré ?? Constructions verbales et accords : signes

vides?, le cas des régimes qui forment aussi des combinaisons liées. Nous distinguons néanmoins le cas des régimes de celui des collocations discuté ici car les régimes constituent des « choix » obligatoires (un lexème ne s'utilise pas sans une construction particulière), alors que le choix d'un collocatif reste toujours optionnel et constitue en cela un véritable choix.

6.12 Fonctions lexicales

Le collocatif étant **choisi en fonction** de la base, on peut modéliser les collocations par des **fonctions**, au sens mathématique du terme, appelées FONCTIONS LEXICALES par Igor Mel'čuk qui les a découvertes à la fin des années 1950 lorsqu'il participait aux premières recherches en traduction automatique en Union Soviétique. À chaque sens qui s'exprime par des collocatifs, on fait correspondre une fonction lexicale qui associe à chaque unité lexicale les collocatifs exprimant ce sens. Le tableau de la section qui précède en est un bon exemple : au sens 'qui n'est pas contrôlée' correspond par exemple une fonction lexicale qui associe panique à PEUR et aveugle à COLÈRE.

On peut voir une fonction lexicale comme une sorte d'unité lexicale dont le signifiant varierait en fonction du contexte, une sorte de super-signème; on aurait ainsi une unité lexicale abstraite signifiant 'qui n'est pas contrôlé' et qui prendrait la valeur *panique* dans un certain contexte, *aveugle* dans un autre ou encore *innocente* en combinaison avec JOIE.

Les collocations n'étant pas prédictibles, elles doivent être listées une à une dans le modèle d'une langue. La liste des collocatifs d'un sémantème s'appelle sa COMBINATOIRE LEXICALE RESTREINTE. La collocation (la co-location en tant que phénomène de semi-figement) s'appelle la COOCCURRENCE LEXICALE RESTREINTE, terme exprimant le fait que la cooccurrence de deux unités lexicales est soumise à des restrictions de sélection.

La combinatoire lexicale restreinte d'une lexie fait partie de son **syntactique**, les autres éléments du syntactique contrôlant sa combinatoire libre et son régime (voir la section 4.4 sur *Signifé*, *signifiant*, *syntactique*). Les collocations constituent ainsi une part très importante de la description du lexique d'une langue. Les dictionnaires monolingues ne font pas une description systématique des collocatifs. Les dictionnaires bilingues en listent généralement davantage du fait qu'un collocatif ne peut pas être traduit indépendamment de sa base. Igor Mel'čuk et son ancien étudiant Alain Polguère ont développé d'importantes bases lexicales pour le français où les collocatifs de chaque entrée lexicale sont décrits à l'aide de fonctions lexicales.



6.13 Collocations morphologiques

Nous avons considéré que les combinaisons propres mais liées étaient des collocations dès qu'elles mettaient en jeu des unités syntaxiques. Ces combinaisons comportent deux unités sémantiques : la base, choisie librement, et le collocatif, dont le choix est contraint par celui de la base.

On est en droit de se demander si les combinaisons propres au sein d'un syntaxème complexe ne sont pas aussi des collocations. Pourquoi des combinaisons très diagrammatiques comme rapide+ment, défend+able, guitar+iste ou fragil+is(er) ne feraient pas aussi l'objet de deux choix : le choix, libre, du radical, puis le choix contraint par le radical du suffixe adéquat.

C'est fort probable que dans certains cas les choses se passent effectivement comme cela. Des erreurs de production, comme le fameux bravitude prononcé par Ségolène Royal alors candidate à la présidence de la république française, tend à prouver que les dérivés de ce type sont parfois construit au moment de la production, avec toutes les erreurs que cela peut entraîner.

6.14 Délimitation paradigmatique des sémantèmes

Nous sommes intéressés jusque-là au découpage de la chaîne parlée en sémantèmes, c'est-à-dire à la délimitation des sémantèmes selon l'axe syntagmatique. Intéressons-nous maintenant au découpage selon l'axe paradigmatique (voir la section 5.10 sur le Signème). Il s'agit de savoir quand deux occurrences d'un même segment de textes appartiennent ou non au même sémantème.

La délimitation des sémantèmes sur l'axe paradigmatique est une question qui dépasse largement le cadre de notre ouvrage et qui est relativement indépendante des questions qui nous intéressent. Nous allons simplement montrer que les principes que nous avons introduits pour la délimitation syntagmatique des unités (principe de commutation, paradigme de choix) permettent aussi une décomposition paradigmatique, ce qui revient à distinguer différentes acceptions d'une même forme.

Nous considérons que deux occurrences de la même forme appartiennent au même sémantème si on peut leur attribuer un même sens, si elles appartiennent au même paradigme de choix (c'est-à-dire que les éléments qui peuvent commuter avec eux soient les mêmes) et si elles ont une combinatoire similaire (et notamment si elles ont les mêmes collocatifs).

Prenons l'exemple du lexème BUREAU. Il possède au moins quatre acceptions bien séparées (que nous numéroterons de 1 à 4 pour les différencier) :

- Dans l'énoncé Le livre est sur le bureau, le choix de BUREAU1 s'oppose à ceux de TABLE ou CHAISE; BUREAU1 dénote un artefact et plus précisément un meuble.
- Dans l'énoncé *Pierre est dans le bureau*, le choix de BUREAU2 s'oppose à ceux de CHAMBRE ou CUISINE; BUREAU2 dénote un lieu et plus précisément une pièce d'habitation.
- Dans l'énoncé *Pierre est au bureau*, le choix de BUREAU3 s'oppose à ceux de MAISON, ÉCOLE ou TOILETTES; BUREAU3 dénote une activité (et se construit avec le verbe support ÊTRE).
- Dans l'énoncé Le bureau s'est rassemblé à nouveau, le choix de BUREAU4 s'oppose à ceux de ÉQUIPE, DIRECTION ou STAFF; BUREAU4 dénote un groupe de personnes.

Comme on le voit, les différents acceptions de Bureau appartiennent à des classes sémantiques assez différentes, même s'il s'agit quand même en un certain sens de la même unité Bureau dont les différents sens sont liés : un bureau2 contient un bureau1, l'activité bureau3 s'exerce dans un bureau2 et un bureau4 est un groupe de personnes qui se rassemblent dans un bureau2. Chaque acception possède sa combinatoire propre et nous les avons d'ailleurs présentées dans des environnements qui permettent de les discriminer : être sur, être dans, être à, se rassembler.

Un autre exemple est fourni par les cinq acceptions suivantes de TOURNER :

- TOURNER1 est lié à TOUR (tourner1 autour de X ~ faire le tour de X) et à CONTOURNER et son dérivé CONTOURNEMENT;
- TOURNER2 est lié à TOURNURE (ça va mal tourner2 ~ ça prend mauvaise tournure);
- TOURNER3 est lié à TOURNAGE (tourner3 un film ~ le tournage d'un film);
- тоurner4 (*le lait tourne*) ne possède aucun dérivé et ni *tour*, ni °*tournement*, ni *tournure*, ni *tournage* ne pourra désigner l'action de tourner du lait.
- тоurner5 (*L'usine tourne à plein régime*) ne possède aucun dérivé non plus et le collocatif *à plein régime* lui est spécifique.

Comme on le voit les dérivations et les collocations ont suffit à elles seules à séparer ces différentes acceptions de TOURNER. Si on reprend l'exemple de PEUR et de ses intensifieurs, on peut distinguer deux acceptions :

- (11) Il a une peur bleue des araignées.
- (12) Tu m'as fais une belle peur. Je ne t'avais pas vu.
- (13) ?? Il a une belle peur des araignées.

?? Tu m'as fais une peur bleue. Je ne t'avais pas vu. (14)

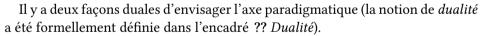
La première acception est une disposition psychique (la peur des araignées), tandis que le deuxième est un sentiment (causé par la surprise). Comme on le voit, ces deux acceptions n'acceptent pas les mêmes intensifieurs, ce qui permet de les distinguer.

Le découpage (selon l'axe paradigmatique) d'un morphème ou d'un syntaxème en sémantèmes donne ce qu'on appelle les acceptions d'une unité.

Les ACCEPTIONS d'un signème sont des sous-ensembles regroupant des signes de même sens et même distribution. Autrement dit, un signème est découpé en acceptions selon les différents sémantèmes dont il est une composante.

Les acceptions d'un syntaxème contiennent aussi des signes désémantisés qui sont des composantes de sémantèmes complexes. Par exemple, on peut ajouter aux acceptions déjà données des lexèmes bureau et tourner leurs occurrences dans les locutions [□]BUREAU DE TABAC [□] OU [□]TOURNER LA PAGE [□].

6.15 L'axe paradigmatique



La première est de considérer l'ensemble des éléments qui peuvent commuter en un point de la chaîne parlée. Autrement dit, on fixe un certain environnement textuel ou structurel qui définit une certaine position structurale et on regarde le paradigme des éléments qui peuvent commuter dans cette position. C'est le point de vue adopté par Saussure lorsqu'il définit les rapports associatifs (voir l'encadré ?? sur La quatrième proportionnelle), même si la définition de Saussure est plus large et inclut aussi des éléments qui ne peuvent pas occuper la même position, mais peuvent se remplacer au travers d'une restructuration complète de l'énoncé (comme durer et pendant dans l'exemple du Chapitre 2 : Zoé a été malade **pendant** deux semaines vs La maladie de Zoé a duré deux semaines).

La deuxième consiste non pas à fixer l'environnement, mais à fixer un élément

Q

et à regarder le paradigme des environnements dans lequel il peut se trouver. C'est ce que nous faisons ici lorsque nous découpons un signème en syntaxème ou en sémantème. Découper en syntaxèmes selon l'axe paradigmatique, c'est regarder les différents environnements possibles d'une forme et découper l'ensemble des signes ayant cette forme pour signifiant selon les environnements dans lesquels ils peuvent apparaître. C'est ainsi qu'on distinguera *avance* lorsqu'il est une occurrence du verbe. Et c'est ainsi que nous définirons au Chapitre ?? les catégories du nom et du verbe.

6.16 Dimension sémiotique du découpage

Nous avons présenté deux axes de découpage, les axes syntagmatiques et paradigmatiques. Il existe un **troisième axe**, que nous appelons l'axe sémiotique, à la suite de Louis **Hjelmslev1943**.

L'AXE SÉMIOTIQUE est la dimension qui relie le sens au texte, le contenu à son expression, le signifié au signifiant.

Parler d'un axe sémiotique, c'est donner une certaine épaisseur à la relation entre le sens et le texte. Le signifié et le signifiant ne se conçoivent pas comme les deux faces d'une feuille, car le signe ne peut pas être seulement considéré dans les relations que ses seuls signifié et signifiant entretiennent avec les signifiés et signifiants d'autres signes. On doit aussi considérer ses relations syntaxiques avec d'autres signes, lesquelles ne sont pas déductibles des relations au niveau des signifiés et des signifiants.

Comme nous l'avons déjà remarqué, les découpages en unités minimales sont différents selon que l'on considère les signifiants, les syntactiques ou les signifiés, ce qui nous donne respectivement les morphèmes, les syntaxèmes et les sémantèmes. On peut voir l'ensemble des productions d'une langue comme un **espace à trois dimensions** : il s'agit d'un fil (la dimension syntagmatique) qui s'enroule sur lui-même à chaque fois qu'il rencontre une nouvelle occurrence d'un signème (la dimension paradigmatique) et qui est composé de strates diverses (la dimension sémiotique). Décrire une langue, c'est d'abord **découper l'espace** tridimension-



nel des productions de cette langue en unités minimales, puis étudier la façon dont ces unités sont liées entre elles.

Prenons l'exemple du morphème vis-dont les principales acceptions sont le nom vis 'tige fileté en hélice', les verbes visser 'mettre une vis', dévisser1 'enlever une vis' et dévisser2 'lâcher prise et tomber' (L'alpiniste a dévissé), le nom VISSAGE 'fait de visser1', l'adjectif VISSÉ (Il est vissé sur sa chaise toute la journée) et la locution SERRER LA VIS 'donner moins de liberté' (Ses parents ont décidé de lui serrer la vis depuis qu'il a des mauvaises notes). Dans vis et visser, le morphème vis- est un syntaxème. Il s'agit de deux syntaxèmes différents, puisque leurs distributions sont totalement différentes. Lorsque *vis-* est une composante de la ^Ilocution ^ISERRER LA VIS^I, il appartient au syntaxème qu'est le nom VIS, dont il est une autre acception; il s'agit d'un signe désémantisé, puisqu'il n'a plus de signifié propre. Le signe DÉVISSER1 est très diagrammatique (dévisser est à visser1 ce que démonter est à monter), de même que vissage (un vissage est à visser1 ce qu'un montage est à monter) : les occurrences du morphème vis- y sont donc des signes, mais ce ne sont plus des syntaxèmes, puisqu'ils n'ont plus de combinatoire libre. Le syntaxème VISSER y a été désyntactisé. Le signe Dé-VISSER2 est peu diagrammatique (il entretient simplement une relation métaphorique avec dévisser1 et le préfixe dé-n'y est pas vraiment motivé, même si l'on a des cohyponymes comme déraper ou dégringoler). Le signe vissé est à peine plus diagrammatique. Les occurrences de vis-dans ces deux derniers sémantèmes sont donc des quasi-signes : le syntaxème visser y a été à la fois désyntactisé et désémantisé.

Il y a donc parmi les sept acceptions considérées du morphème vis-deux quasisignes et cinq signes dont deux sémantèmes (VIS, VISSER), deux lexèmes désyntactisés (lorsqu'il est une composante des sémantèmes DÉVISSER1 et VISSAGE) et un lexème désémantisé (lorsqu'il est une composante de [□]SERRER LA VIS [□]). La figure suivante récapitule cela en montrant les différences entre morphème, syntaxème et sémantème. Chaque ligne correspond à un sémantème, l'acception de vis- étant entourée d'une bulle. Les syntaxèmes sont indiqués par des rectangles pleins et le morphème vis- par un rectangle en pointillé

Figure 6.1 - Découpage paradigmatique du morphème vis-

6.17 Extensions paradigmatique et syntagmatique



(1)

Un signème est un ensemble de signes. On peut mesurer l'extension du signème dans deux dimensions: syntagmatique et paradigmatique. (Voir la section

suivante pour la troisième dimension, la dimension sémiotique.)

L'extension syntagmatique du signème est la portion de texte que couvre le signème, c'est-à-dire l'empan de ses signifiants. Nous notons || <signème> || l'extension syntagmatique d'un signème <signème>. Par exemple, || dévisser ||, l'extension syntagmatique du verbe dévisser, est égale à *déviss-* /devis/. De ce point de vue, les morphèmes sont les signèmes qui ont les extensions syntagmatiques les plus courtes, puisque ce sont les unités de forme minimales. L'extension syntagmatique d'un morphème est toujours contenue dans celle d'un syntaxème (à l'exception des amalgames), qui est elle-même contenue dans celle d'un sémantème. Nous pouvons résumer cela par la formule suivante :

(15)
$$\|$$
 $\|\subseteq\|$ $\|\subseteq\|$ $\|$

(Nous utilisons le signe d'inclusion \subseteq , introduit dans l'encadré ?? sur la *Théorie des ensembles*, car l'extension syntagmatique d'un signème est vue comme une partie du texte.) Par exemple :

(16)
$$||vis-|| \subseteq ||$$
 dévisser $||$ et $||$ vis $|| \subseteq ||$ $||$ $||$ Serrer La vis $||$

L'extension paradigmatique du signème est l'ensemble de ses occurrences possibles, c'est-à-dire l'ensemble des signes et quasi-signes qu'il contient. (Chaque signifié différent du signème donne un signe différent et chaque présence dans un sémantème différent donne un quasi-signe différent.) Nous notons [<signème>] l'extension paradigmatique d'un signème <signème>. Par exemple, [vis-], l'extension paradigmatique du morphème vis-, contient les signifiés 'vis', 'visser1', 'visser2', '\serrer la vis\s\s\s', etc. L'extension paradigmatique d'un sémantème est contenue dans celle d'une combinaison particulière de syntaxèmes, qui est ellemême contenue dans celle d'une combinaison particulière de morphèmes (à l'exception des cas, assez courants, d'allomorphie; voir section 5.20). Nous pouvons résumer cela par la formule suivante :

(17) [
$$<$$
sémantème $>$] \subseteq [$<$ syntaxème $>$] \subseteq [$<$ morphème $>$] Par exemple :

(18)
$$[^{\boxtimes} SERRER LA VIS^{\boxtimes}] \subseteq [VIS] \subseteq [vis-]$$

6.18 Syntaxème et faisceau de signes

Les ensembles de signes qui forment les syntaxèmes possèdent une propriété remarquable qui justifie d'en faire des unités linguistiques à part entière. Il existe

entre les signifiés et les signifiants des signes qui appartiennent à un même syntaxème une forme d'indépendance : chaque association entre un sens particulier et une forme particulière du syntaxème est possible et forme un signe.

Le syntaxème est ainsi un faisceau de signifié et le signifiant, sont **indépendantes** : chaque sens peut se combiner avec n'importe quelle forme et chaque forme avec n'importe quel sens.

Prenons l'exemple du verbe aller qui est à la fois **polymorphique** (/v/, /al/, /i/, /aj/) et **polysémique** ('se déplacer' (*il va quelque part*), 'se sentir' (*comment allezvous*?), 'futur' (*il va partir*), etc., acceptions auxquelles il faut encore ajouter toutes les participations à un phrasème, comme dans [⊠]Aller au Charbon [⊠] ou [™]Aller se faire cuire un œuf D. On peut représenter l'association libre entre les sens et les formes de aller sous la forme d'un faisceau au centre duquel nous plaçons le nom du syntaxème. Comme on le voit, un signe est l'association d'une des formes avec un des sens, tandis que le signème est la réunion de tous ces signes :

signe signème

Figure 6.2

La symétrie entre sens et forme n'est pas complète. Le locuteur fait le choix d'un sens et ce choix peut se porter sur n'importe lequel des signifiés du syntaxème. Mais le « choix » du signifiant n'en est pas un : il est **totalement imposé par l'environnement**, c'est-à-dire par les choix adjacents du locuteur. Dans le cas de Aller, le « choix » du signifiant est conditionné par la flexion et donc par le choix du temps et celui du sujet via l'accord.

Nous montrons pour finir le découpage du syntaxème selon la forme, qui nous donne les (allo)morphes, et le découpage selon le sens, qui nous donne les acceptions.

(allo)morphe acception

Figure 6.3

Il devient alors possible de voir le syntaxème comme un niveau d'articulation

entre sens et forme. On peut ainsi scinder la correspondance sens-forme en deux modules et voir les sémantèmes et les morphèmes comme les éléments respectifs de ces deux modules :

morphe et morphème sémantème

Figure 6.4

En d'autres termes, on peut voir les morphes et morphèmes comme des (ensembles de) demi-signes de surface dont le signifié est un élément abstrait (représenté par le nom de lexème aller dans nos figures), qui sert lui même de signifiant aux demi-signes profonds que sont les sémantèmes. Travailler avec des demi-signes est particulièrement intéressant quand on considère des locutions : on peut en effet considérer que le signifiant de sémantèmes complexes comme Anoir les pieds sur terre ou comme le passé composé sont des configurations de syntaxèmes.

6.19 Signèmes et signes :unités de la langue et unités de la parole Ouverige de la langue et unités de la langue et unités de la parole Ouverige de la langue et unités de la langue et unités de la parole Ouverige de la langue et unités de la langue et unités de la parole Ouverige de la langue et unités de la langue et unités de la langue et unités de la parole Ouverige de la langue et unités de la langue et unités de la parole Ouverige de la langue et unités de la langue et unités de la parole Ouverige de la langue et unités de la langue et unités de la langue et unités de la parole Ouverige de la langue et unités de la langue et unités de la langue et unités de la parole Ouverige de la langue et unités de la langue et unités de la langue et unités de la parole Ouverige de la langue et unités de la langue et unité de la langue e

Depuis les travaux de Saussure, la quasi-totalité des ouvrages de linguistique considèrent que les unités de la langue sont des signes. Ce n'est pas notre cas. Nous considérons que les unités de la langue, celles qui sont dans notre cerveau, sont plutôt des signèmes, c'est-à-dire des **faisceaux de correspondances entre sens et formes**. Lorsqu'un locuteur produit un énoncé, il choisit de tels faisceaux qu'il assemble. Un signème donné est choisi parce que l'un des sens du signème est un sens que le locuteur veut exprimer. Par exemple, si le locuteur veut parler d'un de ses prochains déplacements, il pourra choisir le faisceau ALLER et produire un énoncé comme « *Je vais à Paris la semaine prochaine.* » Le choix de la forme *v*- du faisceau lui est imposé par le contexte : ce sont les sens 'moi' réalisé par *je* et 'présent' qui vont sélectionner l'allomorphe *v*- parmi les différents morphes de ALLER. Ce « choix » du signifiant est totalement indépendant du sens par lequel le locuteur a sélectionné le signème ALLER.

On peut donc dire que ce ne sont pas les signes qui sont organisés en faisceau, mais les faisceaux qui se réalisent par des signes en parole. Lorsqu'un signème est utilisé dans un énoncé, c'est forcément qu'il a été choisi pour un de ses sens et qu'il est réalisé sous une de ses formes. Donc, dans un énoncé, le signème apparaît sous la forme d'un signe. Selon nous, c'est bien un signème qui a été sélectionné dans notre cerveau et c'est par l'intermédiaire de ce faisceau qu'un

signe a été réalisé. En reprenant la terminologie saussurienne, on peut dire que les **signèmes** sont les **unités de la langue** et les **signes** les **unités de la parole**.

Il faut aussi noter que les signes sont quasiment des observables : on peut dans un texte repérer leur signifiant et mesurer leur contribution sémantique. Les signèmes, eux, sont des constructions théoriques. Leur existence ne peut être prouvée qu'au sein du modèle (tant qu'on n'a pas les moyens de les observer au sein du cerveau).

6.20 Autonomie de la syntaxe et de la sémantique

Nous allons conclure ce chapitre en soulignant un point fondamental concernant les unités sémantiques et la syntaxe : le comportement des syntaxèmes (c'est-à-dire leur combinatoire libre avec d'autres syntaxèmes, leur placement ou leur schéma prosodique) dépend assez peu des unités sémantiques auxquelles ils appartiennent. Par exemple, si l'on prend les occurrences de prendre dans n'importe quel phrasème dont il est la tête (Prendre les Jambes à son cou[®], Prendre son pied[®], s'en prendre plein la Gueule[®], etc.), prendre y conserve ses propriété habituelles de verbe, prend une flexion verbale, et peut-être modifié par les mêmes types d'éléments (adverbes, etc.). Ses compléments ne peuvent pas être modifiés (c'est l'effet du figement), mais ils se placent comme des compléments libres et obéissent aux mêmes règles de phonétisation (prosodie comprise).

En général, les phrasèmes ont tendance à avoir une combinatoire plus restreinte et à supporter moins facilement les manipulations que les combinaisons libres, mais ce n'est pas une règle. Par exemple, si on compare le sémantème $^{\boxtimes}$ Briser la glace 'dissiper la gêne' avec la combinaison libre briser \oplus Glace, on voit que les deux combinaisons acceptent d'entrer dans des constructions analogues : négation (*Pierre n'a pas brisé la glace*), clivage du sujet (*C'est Pierre qui a brisé la glace le premier*) ou passif (*La glace a été brisée*). Par contre, le clivage de l'objet (*C'est la glace que Pierre a brisé*) n'est possible qu'avec la combinaison libre, c'est-à-dire quand l'objet est une unité sémantique. Le comportement des phrasèmes dépend donc fortement du comportement des syntaxèmes qui les composent, alors que celui des syntaxèmes est relativement indépendant du fait qu'il forme un sémantème ou sont seulement une composante d'un phrasème.

Morphèmes et sémantèmes structurent le lexique de la langue : ce sont respectivement les plus petites et les plus grandes unités minimales de première articulation. Cependant, le fait que les unités minimales du point de vue du signifiant (les morphèmes) ne correspondent pas aux unités minimales du point de vue du signifié (les sémantèmes) entraine que ni les unes, ni les autres ne sont les unités

de la syntaxe. Nous allons donc pouvoir maintenant présenter les unités de la syntaxe et cette mise au point sur morphèmes et sémantèmes permettra, nous l'espérons, de bien comprendre ce qui est et n'est pas une unité de la syntaxe.



Exercices

Exercice 1 Étudier le caractère motivé ou non des signifiants des signes désignant les nombres de *dix* à *vingt*.

Exercice 2 Alors que la grammaire normative indique que le verbe PAL-LIER doit être utilisé transitivement (*pallier un problème*), de nombreux locuteurs produisent aujourd'hui *pallier à un problème*. Qu'est-ce qui peut expliquer que la construction de ce verbe change de cette façon?

Exercice 3 L'énoncé « Si j'étais toi, je demanderais à **ma** mère. » est ambigu : ma mère peut référer à la mère du locuteur ou à celle de l'interlocuteur. Cela remet-il en question une analyse purement déictique du pronom Moi? Quelle solution proposer?

Exercice 4 a) Montrer que *grièvement blessé* est une collocation.

b) L'adverbe *grièvement* joue un rôle d'intensifieur auprès de *blessé*. Chercher d'autres exemples d'intensifieurs exprimés par un collocatif pour des adjectifs, mais aussi pour des verbes et des noms.

Exercice 5 Identifier les phrasèmes et les collocations des phrases suivantes :

- 1. L'agent de police dormait à poings fermés.
- 2. Il est allé faire la sieste sur la plage.
- 3. Je ne peux pas le piffrer.

4. Il a je ne sais quelle idée en tête.

Exercice 6 Déterminer parmi les occurrences suivantes du lexème TABLE différentes acceptions. Pour chacune d'elle vous indiquerez si elle appartient à un phrasème ou à une collocation.

- 1. Cette table a un pied central.
- 2. On passera à table à midi.
- 3. Tu as intérêt à te mettre à table rapidement si tu ne veux pas te retrouver en tôle.
- 4. Consulter la table des matières en début d'ouvrage.
- 5. On vous a réservé la meilleure table.
- 6. Le résultat est dans la table de la page précédente.

Exercice 7 Un syntaxème comme BLANC peut être utilisé dans la formation de différentes expressions : *voter blanc, blanc comme un linge, blanc comme neige, blanchir le linge, blanchisserie.* Quelles composantes du signe sont affectées dans ces différents emplois?



Lectures additionnelles

Le sémantème, appelé monème par André Martinet, est défini dans ses Éléments de linguistique générale, dont on consultera les sections 1 à 19 du chapitre 1 et le chapitre 4. Les notions introduites par Martinet sont également présentées dans CostaouecGuérin2007. Le chapitre sur Les unités significatives de l'encyclopédie de DucrotSchaeffer1995 complètera cette

lecture.

La syntaxe des **locutions** est étudiée en détail dans la thèse de Marie-Sophie **Pausé2017**. La distinction entre **transparence** et **diagrammaticité** (appelée *analycité*) est bien dégagée dans l'article de Marie Helena **Svensson2008**.

Le découpage paradigmatique en unités lexicales ainsi que les collocations sont très bien présentés dans Polguère2003. Voir également Mel'čuk, ClassPolguère1995 pour une introduction plus technique aux fonctions lexicales, ainsi que Mel'čuk & Polguère2007 pour des exemples d'entrées lexicales avec leurs collocatifs. Les collocations morphologiques sont présentées dans Beck2019.

La notion de **faisceau de signes** est introduite dans le premier tome du *Cours de Morphologie Générale* de **Mel'čuk1993** (voir Chapitre 4) dans la description des grammèmes, sans être exploitée pour les syntaxèmes en général.

David **Beck2019** Phraseology in morphology : It's a sign!, *Linguistica Atlantica*, 37(2), 1-19.

Denis Costaouec & Françoise **Guérin2007** *Syntaxe fonctionnelle - Théorie et exercices*, Presses Universitaires de Rennes.

Oswald Ducrot & Jean-Marie Schaeffer1995 Nouveau dictionnaire encyclopédique des sciences du langage, Seuil.

Louis Trolle **Hjelmslev1946** *Omkring sprogteoriens grundlæggelse*, trad. fr, *Prolégomènes à une théorie du langage*, 1968.

André Martinet (1960, édition revue, corrigée et étendue en 1980), Éléments de linguistique générale, Armand Colin.

Igor Mel'čuk, André Clas, & Alain **Polguère1995** *Introduction à la lexi-cologie explicative et combinatoire*, De Boeck Supérieur.

Igor Mel'čuk & Alain **Polguère2007** *Lexique actif du français : l'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*, De Boeck.

Marie-Sophie **Pausé2017** *Structure lexico-syntaxique des locutions du français et incidences sur leur combinatoire*, thèse de l'Université de Lorraine.

Alain **Polguère2003** *Lexicologie et sémantique lexicale - Notions fonda- mentales.* Presses de l'Université de Montréal.

Maria Helena Svensson2008 A very complex criterion of fixedness : Non-compositionality, in Sylviane Granger & Magali Paquot (dir.) *Phra-*

seology: an interdisciplinary perspective, Benjamins, pages 81–93.



Corrections des exercices

Corrigé 1 Les nombres dix-sept, dix-huit et dix-neuf sont construits selon la syntaxe régulière des nombres (comme cent vingt-huit par exemple). On peut donc considérer qu'ils sont compositionnels. Les nombres onze, douze, treize, quatorze, quinze, seize ne sont pas décomposables en synchronie, même s'ils possèdent tous une terminaison en -ze et un radical en partie transparent. Ils sont donc assez motivés. Les signifants des nombres dix et vingt sont eux totalement arbitraires.

Corrigé 2 Il semble que le verbe pallier ait changé de construction en raison de sa proximité sémantique avec remedier ou $^{\boxtimes}$ s'attaquer $^{\boxtimes}$ qui se construisent avec X V à Y. Le fait qu'un verbe puisse changer de construction montre bien qu'il y a une certaine indépendance de la construction par rapport au lexème lui-même. Voir l'encadré ?? *Constructions verbales et accords : signes vides* ?.

Corrigé 3 Le pronom Moi ne réfère pas forcément au locuteur à proprement parler, mais à celui à qui on attribue les paroles. C'est le cas avec le discours rapporté : *Zoé m'a dit : « Je viendrai ».* C'est aussi le cas ici, où il y a un transfert du 'toi' au 'moi'.

Corrigé 4 L'adverbe Grièvement ne peut modifier que l'adjectif blessé. Blessé garde son sens usuel et Grièvement fonctionne comme un intensifieur. L'absence de commutation possible sur blessé montre que Grièvement est un collocatif. Le choix de blessé est libre et le choix de Grièvement est dépendant de ce premier choix. Les intensifieurs sont les plus productifs des collocatifs. En voici quelques uns : gravement malade, con comme un balai, armé jusqu'au dent, aimer à la folie, applaudir des deux mains, craindre comme la peste, grand ami, gros fumeur, augmentation substantielle, catastrophe épouvantable, désaccord profond, bataille sans merci, forme olympique, etc.

Corrigé 5

- 1. Même si

 AGENT DE POLICE

 est très transparent, il s'agit bel et bien d'un phrasème, et cela pour deux raisons. Premièrement, un agent de police, n'est pas n'importe quelle personne travaillant à la police (ça n'est pas un inspecteur ou un commissaire). Deuxièmement, on peut difficilement faire des commutations sur agent ou police. (Le problème est un peu compliqué ici par le fait qu'il y a une autre acception de AGENT qui à elle seule veut dire 'agent de police'.)

 A POINGS FERMÉS

 est également un phrasème et ce phrasème est un collocatif de DORMIR marquant l'intensification.
- 2. *faire la sieste* est un exemple typique de construction à verbe support, où la base sieste est un nom prédicatif et le collocatif faire n'a pas de contribution propre. Il y a une autre collocation moins évidente qui est *sur la plage*. En effet, la préposition locative sur n'est pas totalement prédictible : on peut dire *à/sur la plage*, mais, par exemple, on doit dire *à/dans la campagne* et seulement *dans la forêt*. On considère donc que les prépositions locatives sont des collocatifs dont le choix est restreint par le nom qui suit.
- 3. Le verbe piffrer ne s'utilise que précédé de POUVOIR (*je ne le piffre pas) et dans un contexte négatif (*je peux le piffrer). Par contre, différentes réalisation de la négation sont possibles : je n'ai jamais pu le piffrer; personne ne peux le piffrer. On a donc un phrasème que nous notons (NE PAS) POUVOIR PIFFRER (la parenthèse indique qu'une autre négation peut remplacer NE PAS).

4. Comme dans les cas précédent, on voit si des commutations sont possibles. On remarque d'abord que le déterminant de IDÉE peut être changé : avoir une idée en tête. De plus, IDÉE peut commuter avec d'autres noms : avoir un problème/une musique en tête. Ou même un pronom : avoir quelque chose en tête. Il nous reste donc l'expression avoir N en tête. Or on peut encore commuter le verbe : on m'a mis cette idée en tête. Au final, on a un phrasème En Tête dont avoir est un collocatif (un verbe support). Quant au déterminant je ne sais quel, il appartient à un paradigme de pronom : je ne sais qui/quoi/où/comment. L'élément je ne sais possède lui-même une certaine variabilité : on ne sait quel, Dieu sait quel. Au final, on considérera quand même que Je ne sais quel est un phrasème, mais cet exemple montre qu'il y a plusieurs niveaux de décomposition et que la frontière entre phrasème et syntagme compositionnel n'est pas parfaitement nette.

Corrigé 6 Le lexème table est un sémantème dans les exemples 1, 5 et 6. Il s'agit de trois acceptions différentes, puisque table désigne un meuble en 1, un lieu gastronomique en 5 et un type de diagramme en 6. En 2, on a une locution $^{\boxtimes}$ À table , comme le montre l'absence de commutation possible sur table. Par contre, passer, qui peut commuter avec $^{\boxtimes}$ SE METTRE $^{\boxtimes}$ ou être, est un collocatif de $^{\boxtimes}$ À table. En 3, la commutation de $^{\boxtimes}$ SE METTRE avec être n'est plus possible et $^{\boxtimes}$ SE METTRE À TABLE forme une locution signifiant 'passer des aveux'. En 4, $^{\boxtimes}$ Table des Matières est également un phrasème.

Corrigé 7 Dans voter blanc, blanc est un collocatif de voter. Dans ce contexte, blanc signifie 'en mettant un bulletin blanc'. Dans blanc comme un linge, c'est comme un linge qui est un collocatif de blanc, mais blanc a une acception particulière 'qui manifeste un sentiment de peur ou de colère'. Par contre, blanc comme neige est un phrasème, car blanc seul ne peut avoir le sens 'sans aucune preuve à charge'; blanc y est donc désémantisé, puisque le sens est porté par la locution complète. Dans blanchir, le radical /blāʃ/ est le même morphème que le syntaxème blanc, mais cette acception n'a plus du tout le même syntactique (il s'agit

d'un conversion de nom en verbe) et son signifié est également modifié et est devenu 'laver à haute température pour rendre blanc et pur'. BLANCHISSERIE est construit à partir de BLANCHIR, mais le signe a subi une nouvelle désyntactisation et désémantisation, puisque même si BLANCHISSERIE est assez diagrammatique, il n'est pas compositionnel pour autant.

Troisième partie

Structures syntaxiques et topologique : Comment les unités se combinent

Présentation

Cette troisième partie est consacrée à la façon dont les signes linguistiques se combinent pour former des syntagmes et aux structures qui en découlent. Le premier chapitre, où nous donnons notre définition de la syntaxe, est consacré à distinguer l'unité minimale de la syntaxe, le syntaxème, du syntagme. Le deuxième chapitre introduit la structure de connexion en montrant comment la combinaison des unités syntaxiques définit un ensemble de connexions qui forment la charpente de la structure syntaxique. Le troisième chapitre montre que cette structure est hiérarchique en introduisant les notions de tête et de dépendance. Le quatrième chapitre étudie les équivalences et différences entre les structure de dépendance, qui mettent en avant les relation syntagmatiques entre unités et les arbres de constituants, qui mettent en avant les relations d'enchâssement entre les unités. Le cinquième chapitre étudie l'ordre des mots et la façon dont les syntaxèmes se regroupent lorsqu'ils sont dans l'ordre linéaire, définissant ainsi ce que nous appelons la structure topologique.

7 Syntaxèmes et syntagmes : La délimitation des unités minimales de la syntaxe

7.1 Syntaxème, morphème et sémantème

Nous avons donné une première définition du syntaxème au Chapitre 6. L'objectif de ce chapitre est de donner une définition plus précise du syntaxème, de préciser la frontière entre le syntaxème et le syntagme, qui est une combinaison de syntaxèmes, et de donner une définition de ce que nous appelons la syntaxe.

Rappelons que nous appelons **morphèmes** les signes minimaux du point de vue de la forme (Chapitre 6) et **sémantèmes** les signes minimaux du point de vue du sens (Chapitre ??). D'où vient la nécessité de considérer, en plus de ces deux unités, les unités que nous appelons les **syntaxèmes**?

Le signe idéal est un sémantème simple, c'est-à-dire une unité minimale de forme et de sens, un sémantème qui est aussi un morphème (Chapitre 6). (Nous devrions dire « un sémantème qui est une acception d'un morphème », car la quasi-totalité des morphèmes ont plusieurs acceptions.) Comme on l'a vu au Chapitre ??, il n'y a pas de correspondance entre unités minimales de sens et unités minimales de forme et un grand nombre de sémantèmes sont complexes, c'est-à-dire sont la combinaison de plusieurs morphèmes. Il convient néanmoins de distinguer parmi les sémantèmes complexes, ceux qui se comportent comme des sémantèmes simples et ceux qui se comportent plutôt comme une combinaison de sémantèmes simples. Ce sont les premiers que nous appelons des syntaxèmes.

Nous avons déjà donné, dans le Chapitre 4, l'exemple des sémantèmes décapsuleur et $^{\boxtimes}$ Avoir les pieds sur terre $^{\boxtimes}$. Les deux sont composés de plusieurs morphèmes, mais, alors que décapsuleur se comporte comme un sémantème simple — comme bol ou couteau —, $^{\boxtimes}$ Avoir les pieds sur terre $^{\boxtimes}$ ne se comporte pas comme un verbe simple. En fait, $^{\boxtimes}$ Avoir les pieds sur terre $^{\boxtimes}$ se comporte comme la combinaison libre *avoir les mains sur la table* et est construit de manière **analogue** à cette expression par la combinaison du verbe Avoir et de compléments avec lesquels il s'est figé sémantiquement. À l'inverse, décapsu-

LEUR n'est pas construit de manière analogue à une expression libre.

7.2 Analogie structurelle

La notion d'analogie (structurelle) joue un rôle central dans la définition des unités syntaxiques. Voici comment nous la définissons :

Une combinaison de signes A+B est dite (STRUCTURELLEMENT) ANALOGUE à une combinaison A'+B' si A a une distribution équivalente à A', B à B' et A+B à A'+B'.

Cette définition va être illustrée dans la section 7.4. Donnons avant cela quelques autres définitions basées sur l'analogie, à commencer par la définition du syntagme :

Un syntagme est une combinaison de signes **structurellement analogue** à une **combinaison libre**.

En incluant les combinaisons *analogues* à des combinaisons libres, nous incluons les constructions figées, c'est-à-dire la possibilité qu'un syntagme ne soit pas une combinaison libre, mais se comporte syntaxiquement comme une combinaison libre.

Les syntagmes sont des combinaisons régulières de signes du point de vue de leur **syntactique** (voir la section 4.4 sur *Signifié*, *syntactique*, *signifiant*). Autrement dit, A+B forme un syntagme si le syntactique de A+B se calcule de manière régulière à partir des syntactiques de A et de B. Mais la combinaison des signifiants de A et B peut être irrégulière (par exemple être un amalgame comme au = a + le ou viens = venirhindicatif+présent+2+singulier) ou la combinaison des signifiés de A et B peut être irrégulière (lorsque A+B est une locution).

La notion de syntaxème peut maintenant être définie de manière plus rigou-

reuse.

Par définition, tout syntagme X est une combinaison A+B qui est libre ou analogue à une combinaison libre. On peut donc décomposer X en deux signes A et B et une telle DÉCOMPOSITION est dite SYNTAXIQUE. Nous opposons une telle décomposition à une DÉCOMPOSITION MORPHOLOGIQUE, qui serait une décomposition en morphèmes reposant uniquement sur la commutation propre.

La décomposition syntaxique peut être appliquée récursivement. Si les signes A et B sont à nouveau des syntagmes, on pourra les décomposer en deux signes et ainsi de suite, jusqu'à ce qu'on arrive à des signes qui ne peuvent plus être décomposés selon ce principe. C'est ce que nous appelons des SYNTAXÈMES.

Toute combinaison de deux syntaxèmes est un syntagme et tout syntagme est une combinaison de plusieurs syntaxèmes.

Plus généralement :

Les unités syntaxiques sont les signes qui commutent librement dans leur contexte ou qui sont analogues à de tels signes.

Une unité syntaxique est soit un syntagme, soit un syntaxème. Et donc :

Les SYNTAXÈMES sont donc les unités syntaxiques minimales. Un syntaxème est indécomposable en deux unités syntaxiques.

7.3 Le collimateur et la sellette



En général, tout syntaxème commute librement dans certains environnements, mais il existe des cas extrêmes comme fur, qui n'apparaît que dans le phrasème \boxtimes AU FUR ET À MESURE. Le fait que *au fur* soit assez clairement analogue à une combinaison libre, en raison notamment de la coordination avec *à mesure*, amène à considérer *fur* comme un syntaxème. Les éléments lexicaux qui ne s'utilisent plus (guère) que dans des phrasèmes sont quand même assez nombreux : (*être*)

aux aguets, (avancer) à la queue leu leu, à l'instar de, (utiliser qqch) à bon/mauvais escient, (poser une question) à brûle-pourpoint, (rouler) à toute blinde, (aller) au diable vauvert, (avoir qqn) dans le collimateur, de plein gré, (coeur) battre la chamade, faire de la charpie (de qqch), de bric et de broc, de guingois, de traviole, en catimini, en filigrane, mettre la sourdine, en un tournemain, en vrac, et tout le bastringue, manger à tous les râteliers, (mettre qqn) sur la sellette, passer au crible, prendre la poudre d'escampette, sans coup férir, sans encombre, se faire du mouron (pour qqn), s'en soucier comme d'une guigne, sonner le glas (de qqch), tailler des croupières (à qqn), (mettre) en exergue, tous azimuts, etc.

7.4 Syntagme ou syntaxème?

Nous allons mettre en pratique nos définitions en étudiant des unités qui sont à la limite entre syntagme et syntaxème.

Commençons en comparant les expressions verbales figées ^図s'en aller et ^図s'enfuir^図. Dans, les deux expressions, le morphème *en* est à l'origine la cliticisation d'un complément délocatif du type de quelque part : fuir de quelque part → en fuir. Il n'y a pas de différence au niveau du figement sémantique entre ces deux expressions. Pourtant elles sont orthographiées différemment et à juste titre. La question est de savoir si les combinaisons en + aller et en + fuir se comportent comme un verbe simple ou comme une combinaison libre clitique en \oplus Verbe du genre en partir (partir de quelque part). Nous allons appliquer la définition de l'analogie structurelle (section 7.2) avec A = A' = en, B = fuir/aller et B' = partir. Il faut chercher les situations où le clitique EN ne se comporte pas de la même façon qu'un préfixe en-. Il n'y en a que deux : la combinaison avec un auxiliaire qui viendra séparer le clitique et le verbe (j'en suis parti) et l'impératif qui inverse l'ordre (Pars-en!). On voit que pour s'enfuir, en- se comporte bien comme un affixe en restant solidaire du radical (je me suis enfui; Enfuis-toi!), mais que $^{\boxtimes}$ s'en aller $^{\boxtimes}$ se comporte encore comme une combinaison en \oplus Verbe (*ie m'en* suis allé; Va-t-en!). On note quand même que la forme je m'en suis allé apparaît comme très soutenue (voire archaïque) et que, dans un style relâché, on pourra avoir [?] je me suis en allé, les locuteurs évitant au final de produire l'une ou l'autre des formes. En conclusion, enfuir est un seul syntaxème, tandis qu'en aller est encore la combinaison de deux syntaxèmes, c'est-à-dire un syntagme.

Un autre exemple est celui de Bonhomme et \boxtimes Bonne femme . Ici, on a deux combinaisons Adjectif + Nom qui se sont figées. Les conventions orthographiques veulent qu'on écrive Bonhomme en un seul mot, bien que pour son pluriel *bons*-



hommes, il soit possible de faire la liaison (/bɔzɔm/), tandis que [⊠]BONNE FEMME s'écrit en deux mots. Voyons comme précédemment si cette différence d'orthographe est bien motivée. Il s'agit donc de savoir si ces deux sémantèmes se comporte comme un nom simple ou comme une combinaison libre Adjectif ⊕ Nom telle que bonne orange. Cela ne sera possible que s'il existe une différence distributionnelle entre le nom et la combinaison Adjectif ⊕ Nom. Une telle différence existe bien : le déterminant indéfini pluriel DES possède une forme faible de qui s'utilise devant un adjectif, tandis que la forme des est obligatoire devant un nom (Pierre a acheté **des** oranges vs Pierre a acheté **de** bonnes oranges). Or l'énoncé [#]Pierre a rencontré de bonnes femmes est impossible avec Bonne FEMME (l'énoncé n'est pas non plus agrammatical, car il est possible avec la combinaison libre BON ⊕ FEMME, mais a alors un autre sens que celui attendu). Le sémantème Bonne femme n'a donc pas la distribution d'une combinaison libre Adjectif ⊕ Nom, mais celle d'un nom simple. Si l'on tient compte de cette différence distributionnelle, il s'agit donc d'un syntaxème. Cette propriété est confirmé par la combinaison avec un lexème comme MINI qui a la propriété de s'accoler au nom : ainsi on peut dire une bonne mini voiture, mais pas *une mini bonne voiture. À l'inverse, on dira sans problème une mini bonne femme.

7.5 À chacun son syntagme



La notion et le terme de syntagme sont empruntés à Saussure qui n'en donne pas de définition formelle. Voici comment il introduit le syntagme :

« Dans le discours, les mots contractent entre eux, en vertu de leur enchaînement, des rapports fondés sur le caractère linéaire de la langue, qui exclut la possibilité de prononcer deux éléments à la fois. Ceux-ci se rangent les uns à la suite des autres sur la chaîne de la parole. Ces combinaisons qui ont pour support l'étendue peuvent être appelées syntagmes. Le syntagme se compose donc toujours de deux ou plusieurs unités consécutives (par exemple: re-lire; contre tous; la vie humaine; Dieu est bon; s'il fait beau temps, nous sortirons, etc.). Placé dans un syntagme, un terme n'acquiert sa valeur que parce qu'il est opposé à ce qui précède ou ce qui suit, ou à tous les deux. » (Saussure1916: 170)

Il est difficile de savoir ce que recouvrait exactement la notion de syntagme dans l'esprit de Saussure. Tout au plus pouvons-nous dire que les exemples donnés ici par Saussure sont compatibles avec notre définition. L'exemple re-lire mérite une discussion, car il est à la limite de ce que nous appelons un syntagme. Le morphème re- est souvent considéré comme un préfixe (ce que laisse notamment supposer la convention orthographique qui le lie au verbe qui suit). Pourtant, à la différence des préfixes usuels, re- se compose très librement avec les verbes, à tel point qu'on est en droit d'en faire un syntaxème. On peut notamment le combiner avec des phrasèmes (?il a renoyé le poisson; ?il a repris le taureau par les cornes), le dupliquer (rerefaire, rerelire) et on trouve des erreurs de production comme Je revais lui dire au lieu de Je vais lui redire, qui laisse penser qu'il est quasiment un clitique préverbal.

L'extrait suivant montre clairement que Saussure inclut dans sa définition du syntagme la combinaison entre un lexème et sa flexion, ce qui est aussi notre cas puisqu'il s'agit d'une combinaison libre (voir Partie 4 pour plus de détails) :

« Quand quelqu'un dit *marchons*!, il pense inconsciemment à divers groupes d'associations à l'intersection desquels se trouve le **syntagme** *marchons*! Celui-ci figure d'une part dans la série *marche*! *marchez*!, et c'est l'opposition de *marchons*! avec ces formes qui détermine le choix; d'autre part, *marchons*! évoque la série *montons*! *mangeons*! etc., au sein de laquelle il est choisi par le même procédé; dans chaque série, on sait ce qu'il faut faire varier pour obtenir la différenciation propre à l'unité cherchée. » (Saussure1916 : 179)

Enfin, contrairement à notre définition, Saussure inclut dans sa définition du syntagme certains faits de morphologie constructionnelle, lorsque ceux-ci relèvent de la parole (c'est-à-dire lorsque les productions font preuve, selon les termes mêmes de Saussure, d'une certaine « liberté de combinaison ») :

« Le propre de la parole, c'est la liberté des combinaisons.

On rencontre d'abord un grand nombre d'expressions qui appartiennent à la langue; ce sont les locutions toutes faites, auxquelles l'usage interdit de rien changer, même si on peut distinguer, à la réflexion, des parties significatives. [...] Mais ce n'est pas tout; il faut attribuer à la langue, non à la parole, tous les types de **syntagmes** construits sur des formes régulières. [...] Quand un mot comme *indécorable* surgit dans la parole, il suppose un type déterminé, et celui-ci à son tour n'est possible que par le souvenir d'un nombre suffisant de mots semblables appartenant à la langue (*impardonnable*, *intolérable*, *infatigable*, etc.). Il en est exactement de même des phrases et des groupes de mots établis sur des patrons réguliers; des combinaisons comme *la terre tourne*, *que vous dit-il*? etc., répondent à des types généraux, qui ont à leur tour leur support dans la langue sous forme de souvenirs concrets.

Mais il faut reconnaître que dans le domaine du **syntagme**, il n'y a pas de limite tranchée entre le fait de langue, marque de l'usage collectif, et le fait de parole, qui dépend de la liberté individuelle. Dans une foule de cas, il est difficile de classer une combinaison d'unités, parce que l'un et l'autre facteurs ont concouru à la produire, et dans des proportions qu'il est impossible de déterminer. »

(Saussure1916: 172)

Notre notion de syntagme, même si elle est plus restrictive que celle de Saussure, couvre davantage de signes que d'autres usages du terme *syntagme* et notamment celui qui est fait par les grammaires dites syntagmatiques (voir Chapitre 10). D'une part, nous considérons que l'unité de base de la syntaxe est le syntaxème et nous appelons bien syntagme toute combinaison de syntaxèmes et non de mots. Ainsi considérons-nous qu'un mot comme *avançait* qui combine librement au moins deux syntaxèmes est un syntagme. Par contre, le signe *indécorable* est selon notre définition un syntaxème complexe, et non un syntagme. D'autre part, nous ne considérons pas, à la différence de l'école anglo-saxonne (qui utilise le terme anglais *phrase*, traduit en français par *syntagme*) qu'un syntagme doive être saturé, c'est-à-dire contenir tous les dépendants de chaque portion du syntagme (cf. Chapitre 9). Ainsi dans *Pierre doit chercher sa montre*, nous considérons que *doit chercher* est un syntagme au même titre que *chercher sa montre* ou *Pierre doit*. Ce point sera largement précisé dans le Chapitre 8 qui suit.

7.6 Syntaxe et morphologie

Nous pouvons maintenant définir ce que nous entendons par syntaxe :

La SYNTAXE est l'étude des combinaisons libres de signes linguistiques et des combinaisons analogues à celles-ci.

Notre définition se distingue des définitions traditionnelles qui voient la syntaxe comme l'étude de l'organisation des mots dans la phrase. Notre définition ne présuppose ni la délimitation préalable d'une unité minimale de la syntaxe (que serait par exemple le mot), ni la délimitation d'une unité maximale de la syntaxe

(que serait la phrase). Notre définition induit une unité minimale de la syntaxe, le syntaxème, que nous définissons en même temps que la syntaxe. La question d'une unité maximale est beaucoup plus complexe et sera abordée dans la partie 6. Notre définition peut être rapprochée de celle d'André Martinet dans son ouvrage de 1985 intitulé *Syntaxe générale*, où la syntaxe est vue comme « l'étude des combinaisons des unités significatives d'une langue » tout en précisant immédiatement que « la syntaxe n'opère pas avec des unités lexicales particulières, mais avec des classes de telles unités » (p. 17), ce qui revient bien à ne considèrer que des combinaisons libres. La définition d'André Martinet néanmoins ne considère pas, contrairement à nous, qu'il puisse y avoir de la syntaxe dans la combinaison des syntaxèmes au sein d'un sémantème, c'est-à-dire dans les combinaisons qui ne sont pas libres, mais seulement analogues à des combinaisons libres.

La **flexion**, qui est de la combinatoire libre de syntaxèmes, relève, avec notre définition, avant tout de la syntaxe : il s'agit d'une composante de la syntaxe que nous appelons la SYNTAXE FLEXIONNELLE, et qui une grande partie de ce que nous appelons la NANOSYNTAXE, qui sera étudiée dans la partie 4. Ceci nous distingue de la grammaire traditionnelle qui considère que l'étude des combinaisons au sein du mot relève de la morphologie. La syntaxe flexionnelle est souvent vue comme relevant à la fois de la morphologie et de la syntaxe; ce qui lui vaut alors le nom de *morphosyntaxe*.

Pour nous, la MORPHOLOGIE est l'étude des formes, c'est-à-dire des signifiants des signes linguistiques. On peut parler de *morphologie flexionnelle*, mais cela ne concerne alors que l'étude des combinaisons des signifiants des syntaxèmes de flexion avec les lexèmes. L'étude de la formation des syntaxèmes est généralement appelée la MORPHOLOGIE CONSTRUCTIONNELLE ou la LEXICOLOGIE, lorsqu'il s'agit de l'étude des lexèmes, c'est-à-dire des syntaxèmes lexicaux.

7.7 Ce que la syntaxe n'est pas

La syntaxe est traditionnellement définie comme « l'étude de l'organisation des mots au sein de la phrase ». Notre définition de la syntaxe s'éloigne de cette définition traditionnelle pour trois raisons.

Premièrement, nous considérons, à la suite de Lucien Tesnière, qu'il existe plusieurs types d'organisation des mots dans la phrase et en particulier une **organisation hiérarchique** et une **organisation linéaire**. C'est la seule organisation hiérarchique que nous appelons **syntaxe**. L'organisation linéaire relève d'une étude différente que nous appelons la **topologie** (Chapitre 11).

Deuxièmement, nous considérons qu'il n'est pas possible de définir le mot



avant de définir la notion de combinaison libre. C'est pourquoi notre définition de la syntaxe repose sur la notion de combinaison libre, plus primitive à notre sens que celle de mot. Il en découle naturellement que l'unité minimale de la syntaxe est défini en termes de combinaison libre : c'est ce que nous avons appelé le syntaxème. Le mot n'est donc pas l'unité minimale de la syntaxe pour nous.

Le syntaxème est une unité plus petite ou égale ou mot (voir néanmoins l'encadré ?? sur les Syntaxèmes séparables). Il s'ensuit que le mot sera défini comme une combinaison de syntaxèmes possédant un niveau de cohésion particulier (Chapitre ??), (Bien que nous ne l'avons pas encore défini formellement, nous nous permettrons de parler de mot tout au long de cette partie, puisque tous nos lecteurs en ont une connaissance intuitive qui est suffisante pour l'instant.)

Troisièmement, nous considérons que la notion de phrase recouvre en fait différentes notions, dont la notion d'unité rectionelle maximale, que l'on ne peut pas définir avant d'avoir défini la notion de rection (voir partie 6).

7.8 Historique de la notion de syntaxe



L'idée que la syntaxe est avant tout l'étude des combinaisons n'est pas nouvelle. Dans sa Grammaire française sur un plan nouveau, avec un Traité de la prononciation des e et un Abrégé des règles de la poésie française publiée en 1709, le jésuite Claude Buffier propose la définition suivante (p. 50) : « La manière de construire un mot avec un autre mot, par rapport à ses diverses terminaisons selon les règles de la Grammaire, s'appelle la syntaxe. » (Nous modernisons l'orthographe.) L'idée de « construction » reste encore très liée à la notion de régime. Le chapitre sur la syntaxe du même ouvrage (p. 294) est intitulé « DE LA SYNTAXE Ou la manière de joindre ensemble les parties d'oraison [= parties du discours] selon leurs divers régimes » et commence par : « Ces diverses parties font pour ainsi dire, par rapport à une langue, ce que font les matériaux par rapport à un édifice : quelque bien préparés qu'ils soient, ils ne feront jamais un palais ou une maison, si on ne les place conformément aux règles de l'architecture. »

Dans l'article « construction » publié en 1754 dans l'Encyclopédie de Diderot et D'Alembert, Dumarsais sépare clairement la syntaxe des questions d'ordre linéaire: « Je crois qu'on ne doit pas confondre construction avec syntaxe. Construction ne présente que l'idée de combinaison et d'arrangement. Cicéron a dit selon trois combinaisons différentes, accepi litteras tuas, tuas accepi litteras, et litteras accepi tuas : il y a là trois constructions, puisqu'il y a trois différents arrangements de mots; cependant il n'y a qu'une syntaxe; car dans chacune de ces constructions il y a les mêmes signes des rapports que les mots ont entre eux, ainsi ces rapports

sont les mêmes dans chacune de ces phrases. »

La stricte séparation de l'ordre structural et de l'ordre linéaire est constitutive de la syntaxe de Lucien Tesnière qui écrit (Tesnières1959, ch. 4-7) : « L'ordre structural des mots est celui selon lequel s'établissent les connexions. [...] Toute la syntaxe structurale repose sur les rapports qui existent entre l'ordre structural et l'ordre linéaire. [...] Parler une langue, c'est en transformer l'ordre structural en ordre linéaire, et inversement comprendre une langue, c'est en transformer l'ordre linéaire en ordre structural. [...] Il y a donc antinomie entre l'ordre structural, qui est à plusieurs dimensions (réduites à deux dans le stemma) et l'ordre linéaire, qui est à une dimension. Cette antinomie est la « quadrature du cercle » du langage. Sa résolution est la condition sine qua non de la parole. »

7.9 Dimension paradigmatique du syntaxème

Comme nous l'avons fait pour les morphèmes, nous regroupons au sein d'un syntaxème différentes occurrences de signes. Pour le morphème, nous avions regroupé tous les signes d'une certaine forme (modulo l'allomorphie, voir la section 5.20) qui possédait une proximité sémantique. Pour le regroupement au sein du syntaxème, nous exigeons que leur distribution syntaxique soit similaire. Notre définition du syntaxème reste de ce point de vue un peu floue : si l'on prend en compte l'ensemble du syntactique, notre regroupement se limitera quasiment uniquement aux signes qui appartiennent à un même sémantème (c'est la position d'Igor Mel'čuk pour qui tout lexème à une acception unique). Nous ne voulons pas pour notre part regrouper uniquement les signes dont le syntactique est absolument identique, mais plutôt ceux dont les syntactiques possèdent un certain recouvrement. Si l'on reprend, l'exemple de avanc-(voir la section 5.10 sur le Signème), nous regrouperons au sein d'un même syntaxème les occurrences qui se combinent avec une flexion verbale et constituent le lexème verbal AVAN-CER et celles qui se combinent avec une flexion nominale et constitue le lexème nominal AVANCE.

Un syntaxème est un signème dont les signes sont minimaux pour la décomposition syntaxique et possèdent un syntactique similaire.

7.10 Typologie des syntaxèmes

Rappelons quelques termes que nous avons introduits au Chapitre 6.

Les syntaxèmes qui appartiennent à des classes ouvertes (verbes, noms, adjectifs, adverbes, interjections) sont appelés des lexèmes. Les syntaxèmes qui appartiennent à des classes fermées indissociables d'un lexème sont appelés des syntaxèmes flexionnels ou grammèmes (voir définition plus précise au Chapitre ??). Les autres syntaxèmes (prépositions, conjonctions, pronoms, déterminants, etc.) sont appelés des lexèmes grammaticaux et forment avec les syntaxèmes flexionnels les syntaxèmes grammaticaux.

Nous appelons syntaxème simple tout syntaxème qui est un emploi libre d'un morphème et syntaxème complexe tout syntaxème multimorphémique.

On ne confondra pas le syntaxème complexe avec le phrasème. Un PHRASÈME est un sémantème qui est aussi un syntagme.

7.11 Constructions N de N



L'exemple le plus emblématique en français de constructions à la frontière de la morphologie et de la syntaxe est certainement la combinaison de deux noms (N) par la préposition *de*.

Tout d'abord, il existe des combinaisons libre N de N comme livre de syntaxe, dont les composantes commutent librement : livre/bouquin/cahier ... de syntaxe/sémantique/géographie ... Or de telles combinaisons sont très cohésives et quasiment inséparables : un livre de syntaxe intéressant vs ^{??} un livre intéressant de syntaxe (voir la section ?? sur la Topologie du groupe substantival en français). Il résulte de cela qu'il n'y a pas de différences de comportement notables entre un N simple et une combinaison libre N de N.

Le problème est donc que si l'on considère un sémantème de la forme N de N, comme $^{\boxtimes}$ POMME DE TERRE $^{\boxtimes}$, il est difficile de dire s'il se comporte comme un N simple (et est un syntaxème complexe) ou s'il se comporte comme une combinaison libre N de N (et est un phrasème). Un tel sémantème n'est absolument pas séparable : une pomme de terre germée vs *une pomme germée de terre. Le plu-

riel n'étant pas prononcé en français (sauf liaison qui ne peut avoir lieu avec de), même si les conventions orthographiques veulent que le -s aille sur pomme, on ne peut pas considérer cela comme une différence avec les noms simples linguistiquement pertinente. On peut contraster cette situation avec celle de l'italien. En italien, les pluriels se prononcent : les noms masculins en -o ont un pluriel en -i. On peut donc s'assurer que le sémantème pomodoro, littéralement pomo d'oro 'pomme d'or', signifiant 'tomate', dont le pluriel est pomodori (et pas pomidoro) est un syntaxème.

La construction N à Vinf est un autre procédé, clairement syntaxique à la base, qui tend aujourd'hui à ne donner que des formes lexicalisées et donc à se morphologiser. La liste des N à Vinf est assez longue, mais ne semble plus permettre de combinaisons libres : poêle à frire, fer à repasser, fer à friser, planche à découper, table à repasser, machine à laver, machine à calculer, graisse à traire, etc. Il reste néanmoins des combinaisons libres du type problème/question/... à résoudre/traiter/reprendre

7.12 Lexème syntagmatique

Il existe une famille de lexèmes qui tout en étant bien des syntaxèmes, c'est-àdire en n'étant analogue à aucune combinaison libre, laisse apparaître une structure syntagmatique figée. Il s'agit de lexèmes comme un lave-linge, un rendezvous, (une idée) à la mors-moi le nœud, un je-m'en-foutiste, je ne sais quelle (idée), il enterre, il atterrit, une bonne femme, parce que, etc. Dans chacun de ces lexèmes, on reconnaît la structure d'un syntagme (lave le linge, je m'en fous, en terre, à terre, etc.), mais le lexème n'est pas analogue à ce syntagme, car il ne possède pas la distribution du syntagme libre. Nous appelons de tels syntaxèmes des lexèmes syntagmatiques (ou des syntagmes lexèmatisés).

Un lexème syntagmatique est une combinaison A+B qui n'est analogue à aucun syntagme, mais qui est la version figée d'un syntagme $A \oplus B$; autrement dit, il existe des combinaisons libres $A' \oplus B'$, où A et A' comme B et B' sont de distributions équivalentes, mais A+B et $A' \oplus B'$ ne le sont pas.

Les signes lexicaux d'un lexème syntagmatique sont très proche de lexèmes,

car il ont conservé leur signifant, une grande partie de leur signifié et une partie de leur syntactique : dans *lave-vaisselle*, la proximité du signe *vaisselle* avec le lexème vaisselle est très importante, puisqu'au niveau sémantique il est bien question de laver la vaisselle et que la combinaison entre *lave* et *vaisselle* s'apparente à la combinaison du verbe avec son objet. Lorsque l'un des éléments correspond à un lexème grammatical comme *en* dans *en+terr(er)*, la combinaison syntaxique devient moins prégnante. Et l'est plus du tout quand la combinaison n'est plus transparente et ne respecte plus la syntaxe du français contemporain comme dans *ce+pendant*.

Il existe également des cas limites de combinaisons totalement atypiques, comme à qui mieux mieux, au petit bonheur la chance ou cucul la praline, qui semblent n'obéir à aucun des procédés de construction de la syntaxe ou de la morphologie ou encore des combinaisons qui attestent de constructions syntaxiques disparues comme l'ordre objet-verbe possible en ancien français : maintenir, ce faisant, tambour battant, sans coup férir, il faut raison garder.

7.13 Constructions N+N ou N⊕N? ♠



Un autre exemple de constructions à la frontière de la morphologie et de la syntaxe est celui des combinaisons N+N de deux noms. D'un côté, il existe des combinaisons fondamentalement liées comme la construction N+N coordonnée, qui associe deux N de manière assez symétrique : un chien-loup, un enseignantchercheur, une moisonneuse-batteuse, un hôtel-restaurant, une fille-mère, un enfantmartyr, etc. Ces combinaisons sont bien liées : si on a un canapé-lit, on n'a pas un °fauteuil-lit ou un °canapé-couchette; si on a la physique-chimie, on n'a pas la °chimie-physique. De l'autre côté, il existe des combinaisons parfaitement libres, comme la **construction** N⊕N **nominative**, qui associe un grand nombre de noms communs avec n'importe quel nom propre : le docteur Mabuse, le général Lee, les frères Coen, l'avenue Victor Hugo, la bibliothèque François Mitterrand, les usines Renault, la station Châtelet, l'affaire Dreyfus, jusqu'à des constructions où le nom propre désigne une époque comme une table Louis XV. Certaines paraissent plus figées : à côté de la région Auvergne, on n'a pas le *département Cantal ou la *ville Saint-Flour. On peut encore rapprocher des constructions nominatives des combinaisons libres similaires comme un bébé phoque ou une mère kangourou.

Entre les deux, on trouve d'autres constructions N+N plus ou moins libres. Commençons par la **construction N+N modificative**, qui apparaît comme la réduction d'une construction N Prép N : *un accès pompiers, une borne incendie, une manif étudiants, un fauteuil relax* (= pour la relaxation), *des pommes vapeur, un*

steak frites, un coin fumeurs. Dans cette construction asymétrique, l'un des deux noms va potentiellement pouvoir se libérer : par exemple, à côté de accès pompiers, on trouvera accès handicapés, accès visiteurs, accès personnel (= du personnel) et donc le nom accès devient ainsi un nom susceptible de régir un N nu, et ce régime est également possible pour des noms sémantiquement similaires comme entrée ou porte. De même des noms comme espace ou coin régissent potentiellement un N nu : coin repas, espace repos, coin fumeurs, espace enfants, coin télé, etc. À noter que cette construction N1\(\theta\)N2 a probablement évoluée d'une construction modificative où le N2 était un modifieur à une construction où le N1 est aujourd'hui recteur. Mais il y aussi des cas où c'est le N2 qui devient productif: par exemple à côté d'accès handicapés, on a aussi un fauteuil handicapés, une rampe handicapés, un ascenseur handicapés, etc. Pourtant même lorsque le N2 devient un modifieur assez libre comme maison dans une confiture maison ou une tarte aux pommes maison, on voit que la construction reste confinée à un N particulier et on n'a pas *une confiture usine ou *une tarte aux pommes boulangerie.

Les constructions N+N coordonnées ne sont pas toujours symétriques : un poisson-chat est un poisson qui a l'allure d'un chat et pas l'inverse. Fonctionnent de manière similaire un requin-marteau, une guerre éclair, une justice escargot ou un discours fleuve. Il nous semble qu'elles relèvent du même procédé de construction que les autres N+N coordonnées, avec pour seule différence un usage métaphorique de N2 : un discours fleuve est un discours au sens propre et un fleuve au sens figuré. Dans les constructions N+N coordonnées asymétriques, N2 peut devenir un modifieur assez libre : satellite/avion espion, personnage/situation clé, maison/grammaire jouet, un livre/film culte/phare/événement.

La frontière entre les différentes constructions N+N n'est pas toujours évidente. On peut penser que c'est le N+N coordonné *stylo-plume* qui va faciliter le passage de *stylo à bille* au N+N modificatif *stylo-bille*. Ou que c'est le N+N nominatif *rayon laser* qui va permettre le passage de *imprimante à laser* au N+N modificatif *imprimante laser* ou au N+N coordonné *sabre-laser*.

Parmi les constructions qui deviennent totalement productives, outre la construction $N \oplus N$ nominative dont nous avons parlé au début, citons la construction qui associe deux aliments, notamment viande \oplus légume, qui est un cas particulier de la construction $N \oplus N$ modificative : une saucisse frites, un steak salade, une truite pommes vapeur, un côte de porc haricots verts. On a aussi un œuf mayonnaise ou un steak sauce au poivre ou même un steak sauce poivre, illustrant la récursivité. Les ingrédients sont des N2 assez productifs; mais si on a une crêpe chocolat ou une gaufre confiture, on n'aura pas *un gâteau pommes ou *une glace

fraise. Par contre avec le doublement de l'ingrédient, on a naturellement une glace vanille-fraise ou même des combinaisons très complexes (et très naturelles) comme une glace deux boules citron vert-chololat amer, où les N modifieurs sont eux-mêmes modifiés. Les matériaux, comme les ingrédients, s'utilisent assez librement comme modifieurs : une peinture métal, une montre or, une toiture ardoise, une finition bois, un revêtement pierre. Comme pour les ingrédients, on ne dira pas *un pantalon coton, mais on aura un pantalon lin-coton.

Que conclure de tout ça? Lorsqu'une combinaison N⊕N est libre, il s'agit soit d'une construction particulière, soit d'un N1 particulier (qui est devenu recteur), soit d'un N2 particulier (qui est devenu un modifieur de nom). Lorsque N2 devient un modifieur complètement libre, comme *une fenêtre standard*, *un cas limite* ou *une chaise marron*, il tend à passer dans la catégorie adjectivale même s'il reste invariable : il peut alors s'employer avec la fonction d'attribut et être modifié par un adverbe (cette fenêtre est parfaitement standard, ce cas est très limite, cette chaise est complètement marron).

En conclusion, la combinaison N+N doit-elle être toujours considérées comme un syntagme? S'il existe indéniablement des combinaisons libres N \oplus N, d'autres combinaisons N+N, comme *chien-loup*, ne le sont pas et il n'est pas certain qu'elles puissent être considérées comme analogues à des combinaisons libres. Dans la mesure où les combinaisons liées N+N sont des constructions avec des sémantiques associées à la combinaison assez différentes des combinaisons N \oplus N, on est en droit de considérer qu'il s'agit dans ce cas d'un procédé morphologique — la composition nominale — et que le résultat est un lexème syntagmatique. Comme pour les constructions N de N, l'absence de différence de comportement entre un N simple et une combinaison libre N \oplus N ne permet pas de trancher de façon définitive le cas des combinaisons liées N+N.

7.14 Construction syntagmatique isolée

Un syntagme est par définition analogue à une combinaison libre. Il existe néanmoins quelques cas de combinaisons de sémantèmes qui ne sont pas analogues à des combinaisons libres, au sens strict où nous avons défini l'analogie, mais que nous voulons néanmoins considérer comme des syntagmes. Le français en offre un bel exemple avec les adjectifs utilisés comme compléments de verbe, comme dans l'énoncé *Cette valise pèse lourd*. Il ne s'agit pas d'une combinaison libre, puisqu'aucun adjectif ne peut commuter avec lourd, même pas léger. Les combinaisons du type PESER + LOURD sont très irrégulières et toujours liées : *coûter cher, parler fort, sonner creux, chanter juste, s'habiller jeune, voter utile,* etc. Il

s'agit donc d'une construction qui n'est directement analogue à aucune combinaison libre du français. Il existe des combinaisons libres Verbe ⊕ Adjectif du type Cette valise paraît lourde, mais elles sont d'un autre type, puisque l'adjectif est un attribut du nom, qu'il s'accorde avec le nom et qu'il commute librement avec d'autres adjectifs. La combinaison PESER + LOURD est pourtant bien un syntagme. Si l'adjectif LOURD ne peut commuter avec un autre adjectif, il peut commuter avec divers groupes nominaux, qui eux commutent librement : Cette valise pèse trente kilos, Cette valise pèse un sacré poids, etc. Il s'agit bien d'une commutation car les deux compléments s'excluent mutuellement : *Cette valise pèse lourd trente kilos. Même si la combinaison peser + lourd est liée, lourd commute proprement avec les groupes nominaux. Le sens de la combinaison PESER + LOURD est compositionnel et il s'agit bien d'une combinaison de deux sémantèmes choisis séparément (même si les deux choix sont liés). Par ailleurs, peser et lourd peuvent être séparés et surtout ils peuvent être modifiés indépendamment l'un de l'autre : Cette valise ne pèse pas lourd, Cette valise pèse plus lourd que prévu. Le comportement de la combinaison PESER + LOURD est donc analogue à celui des combinaisons libres et non à celui des syntaxèmes, qui même lorsqu'ils sont complexes, ne peuvent être modifiés que comme un tout.

Nous ajoutons donc à notre définition du syntagme l'extension suivante :

Toute combinaison de deux sémantèmes est un syntagme, même si cette combinaison est liée et n'est analogue à aucune combinaison libre.

7.15 Inséparabilité linéaire

L'inséparabilité linéaire est certainement la propriété la plus remarquable des syntaxèmes. Elle est souvent utilisée comme propriété définitoire, notamment lorsque les syntaxèmes sont définis à partir des mots.

Une combinaison A+B est (LINÉAIREMENT) SÉPARABLE s'il existe une classe non fermée de syntaxèmes qui peuvent venir s'intercaler entre A et B sans



changer la nature de la combinaison entre A et B.

L'inséparabilité linéaire des syntaxèmes ne peut pas être démontrée (elle peut bien sûr être vérifiée pour les langues déjà décrites, mais pas pour la totalité des syntaxèmes de la totalité des langues). Elle peut simplement être constatée (voir néanmoins les cas limite dans l'encadré ?? qui suit). Elle découle très logiquement du fait qu'il n'y a aucune raison qu'un élément vienne séparer un syntaxème: en effet, pour s'intercaler ainsi dans un segment, il faudrait que l'élément intercalaire dépende d'une seule des deux parties du segment et que donc ce segment soit une combinaison de deux sémantèmes ou se comporte de manière analogue. Ce qui devrait donc nous surprendre n'est pas qu'un syntaxème soit inséparable, mais plutôt qu'un sémantème soit séparable; par exemple, qu'un phrasème comme Noyer le poisson puisse être séparable : dans *Pierre noyait* toujours le poisson, la flexion -ait et l'adverbe toujours se sont intercalés. Ceci montre le caractère particulier des phrasèmes : l'ensemble forme une unité sémantique, mais seule une partie, ici le verbe NOYER, est syntaxiquement accessible et donc les syntaxèmes qui se combinent avec [⊠]NOYER LE POISSON [™] vont en fait se combiner au seul verbe NOYER et se positionner par rapport à lui seul.

C'est le caractère inséparable et l'absence d'indépendance de ses composantes qui rend le syntaxème complexe assez différent d'un phrasème. Il faut néanmoins se souvenir que l'inséparabilité n'est pas un critère pour identifier les syntaxèmes : il s'agit d'un indice soit de figement, soit de cohésion syntaxique. (La cohésion syntaxique et les syntagmes inséparables sont étudiés dans la partie 4 qui suit, consacrée à la nanosyntaxe.)

7.16 Syntaxèmes discontinus



Des langues comme le français pourraient laisser penser que les syntaxèmes se combinent par concaténation, c'est-à-dire s'accolent toujours les uns derrière les autres ou éventuellement se fusionnent (voir la section ?? sur Amalgame, alternance et mégamorphe). Mais, il existe des langues où les syntaxèmes se combinent en s'imbriquant davantage. Tel est le cas des langues sémitiques ou des langues austronésiennes. Arabe. En arabe, la plupart des verbes possèdent un radical formé de 3 consonnes, tandis que la flexion verbale vient compléter ce radical discontinu. Considérons le paradigme :

- (1) a. kataba zajdun 'Zayd a écrit'
 - b. jaktubu zajdun 'Zayd écrit'
 - c. kutiba kitābun 'un livre a été écrit'
 - d. ?akala zajdun 'Zayd a mangé'
 - e. ja?kulu zajdun 'Zayd mange'
 - f. ?ukila tuffāḥatun 'une pomme a été mangée'.

On en déduit que les verbes sont $K_T_B_$ 'écrire' et $?_K_L_$ 'manger', tandis que les amas flexionnels sont $_a_a_a$ = actif.passé.3.sg, $ja__u_u$ = actif.présent.3.sg et $_u_i_a$ = passif.passé.3.sg.

De tels exemples ne remettent pas en cause le caractère inséparable des syntaxèmes. De tels syntaxèmes sont bien **discontinus**, mais les éléments qui viennent s'intercaler entre les différentes parties du lexème sont des syntaxèmes flexionnels, c'est-à-dire des éléments appartenant à une classe fermée et indissociables du lexème. Nous ne parlons de séparabilité que lorsqu'il y a insertion possible d'une classe non fermée d'éléments.

Anglais. Notons encore l'importance de restreindre, dans la définition de la séparabilité, l'élément intercalaire à une classe non fermée. En anglais britannique oral, l'insertion d'un syntaxème comme fucking (a fucking dog 'un putain de chien') est possible à l'intérieur de mots, aux frontières de morphèmes (unfucking-believable 'complètement incroyable') ou bien devant la syllabe accentuée (abso-fucking-lutely 'absolument'). Une telle division de syntaxèmes, appelée une tmèse, est très restreinte lexicalement, et ne met pas en cause l'inséparabilité linéaire du syntaxème.

7.17 Syntaxèmes séparables? 🕏

Certains sémantèmes séparables possèdent une syntaxe suffisamment particulière pour ne pas être analogue à une combinaison libre et être éventuellement déclarés comme des exemples de syntaxèmes séparables. Si de tels éléments existent, ils restent extrêmement marginaux.

L'exemple qui se rapprocherait le plus d'une telle situation est celui des verbes à particule de l'anglais, de l'allemand et des langues germaniques en général.

Verbes à particules (langues germaniques). Les verbes à particules de l'allemand ont été longtemps décrits comme des verbes à préfixe séparable. Nous allons commencer par regarder les verbes à particule de l'anglais avant de présenter ceux de l'allemand.

Les verbes à particule de l'anglais (appelés généralement phrasal verb, littéralement 'verbe syntagmatiques') sont les tournures verbales du type : pick up 'ramasser', go on 'continuer', take off 'enlever', figure out 'arriver à comprendre', etc. Une grande partie de ces tournures sont figées. Si elles l'étaient toutes, ceci fournirait un parfait exemple de syntaxèmes séparables, puisque le complément d'objet peut parfois se placer entre le verbe et la particule : I picked it up 'je l'ai ramassé', they had to take his leg off 'on a dû l'amputer d'une jambe'. Néanmoins, une partie des verbes à particule sont clairement des syntagmes et donc par analogie tous le sont. Il s'agit des combinaisons d'un verbe de déplacement et d'une particule directionnelle : go/walk/swim/crawl/drive/pull/put ... in/out/up/down/through... litt. 'aller/marcher/nager/ramper/conduire/tirer/mettre ... dans/hors/vers le haut/vers le bas/à travers ...' mais qu'on traduirait plutôt en français par entrer/sortir/monter/descendre/traverser ... à pied/à la nage/à quatre pattes/en voiture/en tirant ...). Ici, même si certaines combinaisons peuvent se figer (et devenir ainsi polysémiques) comme go on, toutes les combinaisons sont à peu près possibles et la commutation est bien libre.

Dans la tradition allemande, les verbes du même type sont écrits en un seul mot à l'infinitif : hineingehen, herauslaufen, durchschwimmen, hochziehen, ... Ces formes sont bien pourtant des combinaisons libres de particules (hinein/heraus/hoch/unter/durch ... 'dans / hors / vers le haut / vers le bas / à travers / ...') avec des verbes simples (gehen/laufen/schwimmen/kriechen/fahren/ziehen/stellen ... 'aller/marcher/ nager/ramper/conduire/tirer/mettre ... '). À l'infinitif, la particule précède toujours directement l'infinitif, ce qui pourrait expliquer la convention orthographique. (À noter toutefois que, dans les constructions appelées Zwischenstellung, l'auxiliaire s'interpose entre la particule et le verbe en présence d'un verbe modal comme dans : Er kam, weil er das Boot hoch hat ziehen wollen 'Il est venu parce qu'il a voulu monter le bateau en le tirant'.) Dans les formes finies, l'ordre des mots ressemble plus à celui de l'anglais et la particule peut être très éloignée du lexème verbal : Er zieht das Boot hoch, litt. 'Il tire le bateau vers le haut', c'est-à-dire 'Il hisse le bateau'. (Nous développons dans la section ?? sur la Structure topologique de l'allemand la description du placement du verbe et de la particule en allemand.)

Notons encore que les particules peuvent se figer totalement et devenir alors un vrai préfixe comme dans unterstellen, litt. sous-poser 'laisser entendre', übersetzen, litt. à travers-mettre 'traduire'. Cette construction n'est plus analogue à une combinaison libre, puisque le préfixe n'est plus séparable et il s'agit bien alors d'un lexème (complexe) : Er unterstellt mir ein Rassist zu sein 'Il laisse entendre que je suis raciste' et Er übersetzt den Artikel ins Französische 'Il traduit l'article

en français'. Ces verbes possèdent toujours une acception compositionnelle avec l'ordre habituel : *Er stellt das Auto unter*. litt. 'Il pose la voiture dessous', c'est-à-dire 'Il met la voiture à l'abri'; *Sie setzte mit dem Schiff über*. litt. 'Elle met avec le bateau à travers', c'est-à-dire 'Elle a traversé en bateau'.

Négation double (français). La négation ne ... pas du français est également un candidat intéressant au titre de syntaxème séparable. Le français exprime en effet par deux mots un sens que la majorité des langues expriment par un seul mot; par exemple, Je ne comprends pas se dit No entiendo en espagnol, Ich verstehe nicht en allemand, wŏ bù dŏng en chinois, etc. Pourtant, on considère qu'il s'agit de deux syntaxèmes car ne ou pas peuvent fonctionner seul (Je peux pas venir, Pas de ça ici, Je ne peux accepter ça, Je ne saurais trop vous conseiller de ...) et que ne peut être associé à d'autres éléments, comme plus, personne, rien ou Jamais, ce qui signifie que pas commute librement ici, même si le paradigme est fermé.

Accord. Certains linguistes, comme Zelig Harris ou André Martinet, considèrent des séquences d'accords (comme les trois pluriels dans *les animaux boivent*) comme un seul morphème séparable. Nous avons dit dans l'encadré ?? Constructions verbales et accords : signes vides ? qu'ils nous semblaient effectivement judicieux de considérer la séquence comme l'expression d'un unique sémantème, c'est-à-dire d'un unique choix du locuteur, mais de considérer quand même chaque accord comme un syntaxème flexionnel séparé. Ceci est justifié par le fait qu'il n'y a pas nécessité à ce que les différents syntaxèmes commandés par le sémantème en question soient toujours présents ensemble. Considérer de tels objets comme des syntaxèmes reviendrait non seulement à considérer des syntaxèmes séparables, mais qui plus est des syntaxèmes à géométrie variable possédant selon les contextes un, deux, trois ou davantage encore de morceaux.

Corrélatifs (français). Il existe en français des phrasèmes dit corrélatifs, car composés de deux éléments corrélés et se comportant de manière similaire. Un premier exemple est $^{\boxtimes}$ CI ... $_{\mathbb{C}}A^{\boxtimes}$ comme dans *Fais pas ci*! *Fais pas ça*! ou *un coup comme ci*, *un coup comme ça*. Ce sémantème a la particularité d'imposer le dédoublement de l'élément avec lequel il se combine, chacune de ses parties se combinant avec l'un des deux éléments dédoublés. Comme chacune de ses parties se comporte comme un pronom usuel (par exemple comme ça), on est en droit de considérer qu'il s'agit de deux syntaxèmes et pas d'un seul. Néanmoins les deux parties de $^{\boxtimes}$ CI ... $_{\mathbb{C}}$ CA ne se combinent pas syntaxiquement entre elles et il ne s'agit donc pas non plus d'un syntagme.

Un autre exemple de corrélatif est celui de $^{\boxtimes}$ Plus ... Plus ... $^{\boxtimes}$: *Plus il mange, plus il grossit.* À la différence de $^{\boxtimes}$ CI ... ÇA $^{\boxtimes}$, les deux composantes de $^{\boxtimes}$ Plus ... Plus ... se comportent de manière assez atypiques, même si d'autres adverbes

peuvent venir dans cette position (Quelquefois il mange; Toujours il grossit), ce qui pourrait lui valoir le statut de syntaxème séparable. Ce qui nous fait quand même pencher pour le considérer comme un syntagme est que ses composantes, les deux plus, possèdent encore certaines propriétés du comparatif plus : certes ils n'occupent pas la position du comparatif (*Il mange plus* vs **Plus il mange*), mais ils peuvent commuter avec MOINS (Plus il mange, moins il a d'énergie; Moins il mange, plus il est irritable) et ils acceptent encore des formes supplétives avec certains adjectifs (Plus le vin vieillit, meilleur il est vs Plus le vin vieillit, plus il est bon).



Exercices

Exercice 1 Y a-t-il des combinaisons libres structurellement analogues à haut de gamme dans des chaussures haut de gamme?

Quel problème pose des expressions comme (mener une affaire) tambour battant ou sans coup férir?

Exercice 3 Pour les paires suivantes, discuter s'il vous paraît justifié ou non d'écrire chaque signe en un ou deux mots, sachant qu'un syntaxème s'écrit normalement en un mot, mais pas un syntagme comportant plusieurs lexèmes:

- a. s'en aller vs s'enfuir;
- b. parce que vs puisque;
- c. à côté vs autour;
- d. bonne femme vs bonhomme:
- e. autre chose vs autrefois.

Exercice 4 Est-ce que homme-grenouille est un syntaxème ou un syntagme?

Exercice 5 Est-ce que *il* et *dort* dans *il dort* sont considérés comme linéairement séparable?



Lectures additionnelles

On ne peut pas dire qu'il y ait de consensus actuel sur la question de la frontière entre syntaxe et morphologie. Le fait de considérer la morphologie comme l'étude des formes, et pas seulement l'étude des mots, reprend l'architecture de la Théorie Sens-Texte d'igor Mel'čuk (voir son livre de 1988 ou celui de 2014 avec Jasmina Milićević). Mel'čuk considère néanmoins, à la suite de Tesnière, que le mot est l'unité minimale de la syntaxe, même si dans les représentations syntaxiques qu'il propose les mots sont décomposés en syntaxèmes.

Le livre de Charles Hockett de 1958 reste un des plus bels ouvrages de cette époque. Ici c'est le morphème qui est considéré comme l'unité minimale de l' « analyse grammaticale ». Le mot est défini dans un deuxième temps comme dans notre approche. Hockett ne dégage pas le concept de syntaxème, mais s'en approche en donnant une grande importance aux paradigmes syntaxiques.

Les travaux en grammaire générative et syntaxe X-barre considèrent que la flexion relève de la syntaxe (notamment avec les constituants IP, Inflection Phrase), mais sans aller jusqu'à considérer les syntaxèmes flexionnels comme des signes linguistiques. Les syntaxèmes flexionnels sont traités comme des traits syntaxiques. Voir par exemple l'article de Stephen Anderson1982.

La morphologie lexématique est présentée dans le traité de morphologie de Bernard Fradin2003. Les combinaisons N+N y sont étudiées (p. 201).

Les combinaisons figées Verbe-Adjectif, comme peser lourd ou sonner creux sont étudiées par LeGoffic1993.

Pour ceux qui s'intéressent à la négation double, on pourra consulter le travail de Jespersen1917 et les travaux ultérieurs qui y font référence en tant que cycle de Jespersen : au cours de l'évolution d'une langue, la négation tend à s'affaiblir, puis à être doublé par un autre morphème qui fini par supplanter le morphème initial et ainsi de suite.

Anderson Stephen R. Anderson1982 Where's morphology?, Linguistic Inquiry, 13, 571-612

Fradin Bernard2003 Nouvelles approches en morphologie. Presses universitaires de France.

Hockett Charles F. Hockett1958 A Course in Modern Linguistics. The Macmillan Company: New York.

Jespersen Otto1917 Negation in English and other languages, Copenhague.

Le Goffic **Pierre1993** *Grammaire de la phrase française*, Hachette.

Mel'čuk **Igor1993** *Cours de morphologie générale*, volume1 (Le mot), Presses de l'Université de Montréal.

Mel'čuk Igor, Milićević Jasmina2014 Introduction à la linguistique, volume 3, Hermann.



Corrections des exercices

Corrigé 1 Dans haut de gamme, haut est un nom (le haut de la gamme). Il s'agit donc d'une construction « N de N » qui a la distribution d'un adjectif. Il n'existe pas de combinaison libre de ce type en français. Ceci

nous amène à considérer *haut de gamme* dans l'emploi d'adjectif comme un **lexème syntagmatique** (voir section 7.12 éponyme).

Corrigé 2 Les expressions *tambour battant* et *sans coup férir* datent d'une époque où les compléments pouvaient être placés devant le verbe. Si ces constructions sont clairement d'origine syntaxique, elles ne sont plus aujourd'hui analogues à aucune construction libre du français et doivent donc être considérées comme des lexèmes syntagmatiques (voir section 7.12 éponyme).

Corrigé 3 a. enfuir est un lexème syntagmatique, mais en aller a encore des propriétés syntaxiques (voir section 7.4 Syntagme ou syntaxème? pour les détails).

b. parce que est l'exemple type d'un lexème syntagmatique dont l'orthographe n'est pas motivé. L'origine de l'expression est par ce que et parce n'est pas un lexème du français. À l'inverse, puisque est dans le paradigme de alors que, bien que ou lorsque où un adverbe se combine avec la conjonction que. Néanmoins, la très faible diagrammaticité de la combinaison puis + que et la prononciation particulière de puis /puis/ dans puisque justifie tout particulièrement de ne pas décomposer en synchronie.

- c. Les combinaisons entre une préposition et un nom sans déterminant, comme \grave{a} $c\^{o}t\acute{e}$, sont \grave{a} la limite entre syntagme et lexème syntagmatique, car d'un côté la combinaison entre une préposition et un nom est généralement libre, mais il n'existe pas de combinaison libre avec la préposition \grave{a} lorsque le nom est sans déterminant. Dans le cas autour, il s'agit plus clairement d'un syntagme, puisque le déterminant est présent et donc la convention orthographique est peu motivée.
- d. bonhomme et bonne femme sont tous les deux des lexèmes syntagmatiques (voir section 7.4 Syntagme ou syntaxème?).
- e. On est ici face à des constructions qui sont clairement d'origine syntaxique, mais qui sont devenues irrégulières aujourd'hui en raison de l'absence d'article. Il est néanmoins clair que la combinatoire de *autre fois* est déterminée par le nom temporel *fois* (voir le Chapitre ??), ce qui justifierait de l'orthographier *autre fois*.



Corrigé 4 Comme nous avons pu le voir dans la section ?? Constructions N+N ou $N\oplus N$?, les N+N coordonnés asymétriques peuvent parfois être libres (satellite espion, film culte), mais cela reste suffisamment marginal pour qu'on considère plutôt homme-grenouille comme un lexème syntagmatique.

Corrigé 5 Les seuls syntaxèmes qui peuvent s'intercaler entre il et dort sont des clitiques (il n'y dort pas). Il n'y donc pas de classe ouverte de syntaxèmes qui peut venir séparer ces deux mots, qui sont donc considérés comme linéairement inséparables, selon notre définition donnée dans la section 7.15 Inséparabilité linéaire.

8 La connexion syntaxique : Décrire les combinaisons entre unités syntaxiques

8.1 Structure syntaxique

Nous savons que les signes linguistiques se combinent pour former des signes linguistiques plus étendus. Nous souhaitons maintenant comprendre comment s'organisent les combinaisons entre signes et quelle structure forme l'ensemble des combinaisons. Une telle structure est ce que nous appelons une STRUCTURE SYNTAXIQUE.

Nous concevons la structure syntaxique comme un **objet mathématique**, une structure au sens mathématique du terme (voir le Chapitre 3 sur *La modélisation*). En représentant la façon dont les signes linguistiques se combinent, la structure permet de mettre en évidence les principales **contraintes** qui régissent ces combinaisons. La structure sert donc de base à l'écriture d'un modèle de la grammaire de la langue.

Notre conception de ce qu'est une structure syntaxique est axiomatique. La langue obéit à un grand nombre de contraintes syntaxiques, dont une partie reste à découvrir. Lorsque nous proposons *une* structure syntaxique, c'est toujours une modélisation partielle de ces contraintes, qui fonctionnent comme autant d'axiomes de base de la modélisation. On peut pour différentes raisons réduire volontairement le nombre de contraintes prises en compte, notamment pour des raisons pédagogiques ou pratiques. Nous n'hésiterons donc pas par la suite à présenter différentes représentations syntaxiques d'un même énoncé. (Voir notamment la section 8.18 sur *Structures de connexion*, *granularité et critères*.)

Dans ce chapitre, nous présenterons une première structure syntaxique qui rend essentiellement compte d'une contrainte, la possibilité pour une portion de l'énoncé de former une **unité autonome**. Nous appellerons cette structure la **structure de connexion**. Dans le Chapitre 9, nous prendrons en compte des contraintes supplémentaires et définirons la **structure de dépendance**.

Insistons aussi sur le fait qu'une structure syntaxique modélise des contraintes

syntaxiques qui apparaissent à l'observation des énoncés qu'un locuteur a produit et peut produire, mais ne découlent, en aucun cas, de l'observation du mécanisme de production en tant que tel, lequel est enfoui dans le cerveau et encore en grande partie inobservable aujourd'hui. Si tout un chacun est autorisé à penser que la structuration syntaxique est une propriété fondamentale de la langue vue comme correspondance entre sens et textes, ce n'est dans l'état actuel des connaissances qu'une hypothèse à vérifier.

8.2 Justifier la structure syntaxique

Rappelons le type d'arguments que nous avons introduits dans le Chapitre 2 pour justifier l'introduction d'une structure syntaxique. Nous avons montré qu'il existait des contraintes sur la production d'un énoncé et qu'on pouvait à juste titre considérer que ces contraintes s'appliquaient de façon hiérarchique, chaque choix à son tour imposant des contraintes sur les choix suivants. Donnons un nouvel exemple :

- (1) La France produit du lait.
- (2) la production laitière de la France
- (3) la production française de lait

Ces trois syntagmes expriment les mêmes sens avec les mêmes relations sémantiques : un prédicat binaire, 'produire', exprimable par le verbe PRODUIRE ou le nom PRODUCTION, et deux éléments de sens, 'la France' et 'lait', qui remplissent les rôles d'agent (celui qui produit) et de patient (ce qui est produit) de ce prédicat, ce qu'on peut schématiser comme suit (voir le Chapitre ?? pour plus de détails) :

Figure 8.1 – Représentation sémantique

Le sens 'la France' est réalisé par un groupe substantival sujet quand 'produire' est réalisé par un verbe (*la France produit*). (Nous appelons *groupe substantival* ce que l'on nomme traditionnellement un groupe nominal. La distinction que nous faisons entre substantif et nom sera justifiée dans la section ?? sur *Substantifs et noms*.). Quand 'produire' est réalisé par un nom, alors les choses se passent différemment : 'la France' peut être réalisé par un adjectif (*la production française*) et s'il est réalisé par un groupe substantival, il doit être précédé de



la préposition DE (la production de la France). C'est cela que nous entendons par des contraintes syntaxiques : les catégories des éléments lexicaux (verbe, nom, adjectif, ...) qui se combinent sont corrélées, les choix lexicaux d'éléments liés entre eux se contraignent les uns les autres.

8.3 Faut-il définir la structure syntaxique?



Ce chapitre et les suivants sont consacrés à la définition de la structure syntaxique. Il est néanmoins légitime de se demander si la structure syntaxique doit être définie et si elle doit l'être par une méthode déductive comme nous allons le faire.

Un linguiste comme Lucien Tesnière est un mentaliste. Pour lui, la structure syntaxique est un produit de l'esprit (« l'esprit aperçoit des connexions », voir citation plus complète à la section 8.8). La définition de la structure syntaxique relève donc de l'observation de ce qu'il y a dans la tête du locuteur. Le problème est que nous ne voyons pas dans la tête des locuteurs et que même les plus fins des outils d'imagerie cérébrale d'aujourd'hui ne permettent pas d'observer s'il y a des arbres syntaxiques dans la tête des locuteurs lorsqu'ils produisent ou entendent un énoncé. Certains linguistes vous diront qu'ils voient les arbres dans leur tête quand ils parlent et que si vous étiez un aussi bon linguiste qu'eux vous les verriez aussi (nous ne donnerons pas de noms & [FE0E?]). On peut leur répondre que c'est probablement le linguiste en eux qui voit des arbres et non le locuteur et que le fait qu'il pense voir des arbres quand il parle ne prouve pas grand-chose.

A contrario, notre méthode n'est pas basée sur l'observation de ce qu'il y a dans la tête des locuteurs, mais sur l'observation des énoncés et de leurs propriétés. Notre méthode est déductive dans le sens où nous déduisons la structure syntaxique d'un énoncé à partir d'un certain nombre de propriétés de l'énoncé que nous observons. Chacune des propriétés considérées doit être caractérisée par un certain nombre de critères à remplir. Si un énoncé ou une portion donnée de l'énoncé ne remplit pas les critères demandés, alors la propriété n'est pas vérifiée. Ainsi chaque propriété peut être falsifiée et la construction de la structure syntaxique peut être reproduite par un autre linguiste (voir le Chapitre 3 sur La modélisation).

Il est à noter que les mentalistes aussi énoncent des propriétés des structures qu'ils considèrent. Mais ces propriétés ne servent pas à définir les structures qu'ils proposent. Tout se passe comme si les structures préexistaient et qu'ils en observaient les propriétés. Autrement dit, s'il advient que l'une des structures qu'ils proposent ne vérifient pas les propriétés énoncées, ce n'est pas la structure qui en est défaut, ce sont les propriétés qui sont à revoir. Ainsi, pour chaque nouvelle construction, une structure d'un type différent peut être proposée et les propriétés générales des structures peuvent être revues. En un sens, nous aussi nous sommes susceptibles face à une nouvelle construction de proposer une structure d'un type différent. Mais cela sera justifié par le fait que la nouvelle construction étudiée présente des propriétés jamais rencontrées jusque-là et que pour en rendre compte il est nécessaire d'introduire un nouveau type d'éléments dans la structure syntaxique.

Il existe une troisième façon de définir la structure syntaxique, qui est celle, nous semble-t-il, défendue par Noam Chomsky: la structure syntaxique n'a pas à être justifiée a priori. Elle l'est a posteriori, par l'écriture d'une grammaire formelle. C'est la possibilité de pouvoir écrire une grammaire basée sur les structures considérées et qui génère toutes les phrases de la langues qui justifie le bien fondé de la structure retenue, quand ce n'est pas la grammaire formelle ellemême qui génère les structures syntaxiques. Ainsi les ouvrages de GRAMMAIRE GÉNÉRATIVE ne justifient-ils jamais vraiment le fait que la structure syntaxique soit représentée par un arbre de constituants, comme si cela ne relevait pas d'un choix du linguiste modélisateur. Il serait pourtant naïf de penser que c'est le formalisme qui décide de lui-même le type de structures qu'il produit. C'est bien là un choix fait a priori par le linguiste lorsqu'il construit ou choisit le formalisme dans laquelle il écrira sa grammaire.

Malgré cette dernière remarque, nous ne sommes pas en réel désaccord avec la dernière approche présentée. D'une part, le choix de la structure dépend évidemment de la grammaire que l'on souhaite écrire. En un sens, définir une structure syntaxique qui rende compte d'un certain nombre de propriétés, c'est se garantir que la grammaire que nous écrirons rendra bien compte de ces propriétés. Les deux points de vue ne se contredisent donc pas. D'autre part, certains choix de structure apparaissent finalement sans importance lorsqu'on écrit la grammaire formelle et c'est l'économie du système (c'est-à-dire la possibilité d'écrire une grammaire plus simple) qui permettra de décider entre deux options de représentation. De ce point de vue, le choix de la structure perd de son importance, dans la mesure où elle n'est qu'un support à la véritable modélisation qui est effectuée par l'écriture d'une grammaire.

8.4 Nature de la structure

Il a été souvent postulé que la structure syntaxique est un **arbre**, arbre de constituants (voir Chapitre 10) ou arbre de dépendance (voir Chapitre 9). L'arbre



est la plus simple des structures hiérarchiques : dans un arbre chaque élément de la structure, chaque nœud pour être plus précis, dépend d'un autre nœud à l'exception d'un seul qui domine tous les autres et qu'on appelle la racine (voir la section ?? sur Graphe et arbre). La plupart, si ce n'est la totalité, des travaux qui postulent que la structure syntaxique est un arbre le justifie par des propriétés de la structure d'arbre et pas par des propriétés de la syntaxe des langues. Si l'on présuppose que la structure syntaxique est un arbre et que l'on fixe des critères pour définir un arbre, on obtiendra forcément un arbre. Ceci ne prouve pas pour autant que la structure syntaxique soit bien un arbre, ni même qu'il existe réellement une structure simple qui représente l'organisation syntaxique des signes.

Dans la suite, nous éviterons tout présupposé sur la nature générale de la structure. Nous verrons que différents principes de modélisation amènent à des résultats différents. Et que si certaines sont naturellement proches d'un arbre, certains indices laissent à penser que la structure est plus complexe. Si la structure est globalement hiérarchique (comme nous l'avons défendu dans le Chapitre 2 et comme nous le défendrons avec plus de détails dans le Chapitre 9), elle n'est pas nécessairement complètement hiérarchique et nous soutiendrons que certaines constructions peuvent ne pas l'être. De plus, supposer que la structure est un arbre, c'est supposer en particulier que les relations se font nécessairement deux à deux, ce qui ne va pas toujours de soi. Enfin, même si une structure d'arbre était avérée, celle-ci ne suffirait pas à encoder tous les types de relations syntaxiques qui existent et la plupart des approches utilisent des moyens additionnels qui ajoutent plus ou moins explicitement de la structure à l'arbre : structure complexe des nœuds de l'arbre, structure des étiquettes catégorielles des nœuds, liens additionnels entre nœuds (par exemple par la co-indexation des nœuds), etc. Nous essayerons pour notre part de rendre explicite la totalité de la structure.

8.5 Calculer la structure syntaxique d'un énoncé



Le fait d'associer une structure syntaxique à un énoncé est une préoccupation ancienne, que l'on peut probablement faire remonter aux tous premiers linguistes et notamment au grammairien indien antique Pānini. Saussure, le père de la linguistique contemporaine, n'utilise pourtant pas le terme structure dans cette acception précise. Les structuralistes, qui le suivent, s'intéresseront davantage à la structure de la langue en tant que système qu'à la structure particulière des énoncés. Cette différence de point de vue est bien exposée par Émile Benveniste :

« Quand les linguistes ont commencé, à l'instar de F. de Saussure, à envisager la langue en elle-même et pour elle-même, ils ont reconnu ce principe qui allait devenir le principe fondamental de la linguistique moderne, que la langue forme un système. Ceci vaut pour toute langue, quelle que soit la culture où elle est en usage, à quelque état historique que nous la prenions. De la base au sommet, depuis les sons jusqu'aux formes d'expression les plus complexes, la langue est un arrangement systématique de parties. Elle se compose d'éléments formels articulés en combinaisons variables, d'après certains principes de structure. Voilà le second terme clé de la linguistique, la structure. On entend d'abord par là la structure du système linguistique, dévoilée progressivement à partir de cette observation qu'une langue ne comporte jamais qu'un nombre réduit d'éléments de base, mais que ces éléments, peu nombreux en eux-mêmes, se prêtent à un grand nombre de combinaisons. On ne les atteint même qu'au sein de ces combinaisons. Or l'analyse méthodique conduit à reconnaître qu'une langue ne retient jamais qu'une petite partie des combinaisons, fort nombreuses en théorie, qui résulteraient de ces éléments minimaux assemblés. Cette restriction dessine certaines configurations spécifiques, variables selon les systèmes linguistiques envisagés. C'est là d'abord ce qu'on entend par structure : des types particuliers de relations articulant les unités d'un certain niveau. » (Émile Benveniste, 1966, Problèmes de linguistique générale, vol. 1, chap. 2, Coup d'œil sur le développement de la linguistique, p. 21, nous soulignons)

Les restrictions qu'une langue impose à la combinatoire des éléments minimaux dont Benveniste parle ici, ne sont pas de nature sémantique (une combinaison qui ne fait pas de sens), mais de nature structurelle : c'est la langue en elle-même, pas le monde dont on parle, qui impose et interdit certaines structures.

On trouve bien avant les travaux de Saussure des descriptions détaillées de structures linguistiques, notamment dans les grammaires destinées à l'enseignement des langues. Nous en discutons dans l'encadré ?? où nous présenterons un Historique des notions de dépendance et de tête. Si l'on trouve de nombreuses descriptions détaillées de structures syntaxiques dans les grammaires du 18^e siècle, il semble que ce ne soit qu'au milieu du 19^e siècle que les linguistes commencent à développer des modèles de représentation de la structure syntaxique à l'aide de diagrammes. Ces premiers modèles, comme ceux pour l'américain de Frederick A. P. Barnard de 1836, pour l'anglais de Stephen W. Clark de 1847 ou pour l'allemand de Franz Kern de 1883, sont développés à des fins pédagogiques pour décrire diverses constructions d'une langue particulière (nous y reviendrons dans l'encadré ?? consacré à l'Historique des représentations syntaxiques par des dia-

grammes en dépendance et l'encadré ?? consacré à l'Historique des représentations syntaxiques par des diagrammes en constituants). Il faudra attendre le 20^e siècle pour que des modèles similaires servent de bases à des théories de linguistique générale appliquées à plusieurs langues, comme dans les travaux de Lucien Tesnière.

Parallèlement, s'est développée, à partir des travaux du philosophe polonais Kaziemierz Ajdukiewicz (un article fameux de 1935 intitulé *Die syntaktische Konnexität*), puis de Yehoshua Bar-Hillel et Zelig Harris dans les années 1950, l'idée qu'on pouvait écrire un système de règles formelles permettant de vérifier qu'une suite de mots est un énoncé bien formé (voir l'encadré ?? *Calcul symbolique et grammaires catégorielles*). Les deux courants vont converger avec les travaux de Noam Chomsky (*Syntactic structures*, 1957), qui le premier considère qu'un énoncé est bien formé si on peut lui associer une structure syntaxique et propose un système de règles pour calculer de telles structures. Les travaux de Chomsky vont avoir une influence considérable. Néanmoins nous pensons qu'ils ont aussi entraîné une grande partie de la communauté linguistique dans une mauvaise direction en se basant trop fortement sur une représentation de la structure syntaxique insatisfaisante (l'analyse en constituants immédiats, voir section ??), en prenant trop tardivement en compte la représentation du sens et en négligeant l'importance des idiosyncrasies lexicales.

8.6 Structure syntaxique et structure topologique

Les travaux de Lucien Tesnière ont mis en évidence deux niveaux d'organisation, qu'il appelle l'ordre linéaire et l'ordre structural. L'ORDRE LINÉAIRE, à prendre ici au sens d'organisation séquentielle, est le fait que les mots se suivent les uns les autres, formant une « ligne ». L'ordre linéaire est donc cette relation de contiguïté et précédence qui existe entre les mots sur l'axe syntagmatique :

- les mots ne se chevauchent pas (SEGMENTABILITÉ);
- les mots sont à côté d'autres mots (CONTIGUÏTÉ);
- les mots sont avant d'autres mots (PRÉCÉDENCE).

La segmentabilité ne va pas tout a fait de soi : elle revient à considérer que lorsqu'il y a amalgame de deux signes, comme dans *au* (à+le) ou *du* (de+le), il n'y a bien qu'un mot. (Voir aussi l'encadré ?? sur *Les exceptions à la segmentabilité*.) La contiguïté, comme la précédence, suppose la segmentabilité (quand deux éléments sont au même endroit, nous pouvons parler d'amalgame ou de chevauchement, mais pas de contiguïté). Par contre, contiguïté et précédence sont des notions indépendantes. Dans *Lise a fini, Lise* précède *fini*, mais n'est pas contigu.

La combinaison de la précédence et de la contiguïté s'appelle la PRÉCÉDENCE IM-MÉDIATE.

Être contigus n'implique pas qu'on soit en relation « structurale ». Prenons un exemple:

Le chat de Marie dort sur le canapé.

Les mots Marie et dort se succèdent, mais n'entretiennent pas de relation sémantique directe : c'est le chat qui dort et non Marie. D'un point de vue syntaxique, le mot Marie ne se combine donc pas avec le mot dort, mais avec le chat via la préposition de, exprimant que le chat appartient à Marie. C'est le résultat de cette combinaison, le chat de Marie, si ce n'est le chat seul, qui se combine avec dort. Ce sont les combinaisons à ce niveau que veut saisir l'ORDRE STRUCTURAL.

Dans la suite, nous utiliserons le terme STRUCTURE SYNTAXIQUE pour parler de l'ordre structural au sens de Tesnière, c'est-à-dire la facon dont les signes se combinent indépendamment de l'ordre linéaire. Et nous appellerons structure topologique la structure obtenue lorsqu'on analyse la relation entre l'ordre structural et l'ordre linéaire.

La structure topologique sera présentée au Chapitre 11.

8.7 De la non-séparation des ordres au mouvement **Q**



L'approche dominante au 20^e siècle, l'analyse en constituants immédiats (voir section ??), inspirée par les travaux de Bloomfield, des distributionnalistes, puis de Chomsky, ne repose pas sur une séparation préalable des ordres linéaire et structural. La combinaison des signes est envisagée comme étant à la fois linéaire et structurale. Une telle approche mène immanquablement à postuler des positions « vides » et des mouvements. Expliquons-nous à partir d'un exemple :

À qui veux-tu que je parle?

Il est clair que à qui se combine avec je parle, de la même façon que à Marie se combine avec je parle dans Tu veux que je parle à Marie : c'est le verbe PARLER (et non le verbe vouloir) qui permet la présence d'un complément du type à Marie ou à qui. Le problème est que à qui n'est pas contigu à je parle : comment rendre compte alors du fait que ces deux éléments se combinent si l'on ne s'autorise pas à considérer des relations indépendamment de l'ordre linéaire? Une solution a été proposée dans les années 1960 : à qui se combine avec je parle avant de se déplacer en tête de la phrase. Selon Chomsky, il y a donc un моиvемент qui s'opère à l'intérieur de la structure d'une position syntaxique vers une autre. (Une position syntaxique est une position dans la structure syntaxique peu importe comment celle-ci est définie; dans le cas de l'analyse chomskienne, il s'agit d'un nœud de l'arbre de constituants.) La position initiale de à qui, celle qu'il occupait quand il s'est combiné avec je parle, est laissée vide : c'est une position vide, où ne subsiste tout au plus qu'une TRACE laissée par l'élément déplacé. La relation entre les deux positions est indiquée par une CO-INDEXATION: chaque position syntaxique reçoit un indice et deux positions qui correspondent au même élément ont un indice identique. Cette analyse est généralement représentée comme ci-dessous, ε représentant la trace laissée dans la position vide et i l'indice commun aux deux positions en relation:

(6) [À qui]_i veux-tu que [je parle
$$\varepsilon_i$$
]?

De nombreux phénomènes syntaxiques ont été analysés en ces termes dans les années 1960 et 1970 et la terminologie utilisée encore aujourd'hui s'en ressent. Ainsi le phénomène mis en évidence par la phrase À qui veux-tu que je parle? est appelé une extraction : à qui a été « extrait » de la proposition subordonnée. Dans une phrase telle Pierre lui a parlé, on parle de montée du clitique (angl. clitic climbing), car le clitique lui « monte » sur l'auxiliaire. Dans une phrase comme Mary gave him a book 'Marie lui a donné un livre', on parle de dativeshift, car le complément au datif (le to him 'à lui' de la supposée source Mary gave a book to him) « change » de place (ang. to shift). Enfin quand l'ordre devient trop complexe pour être décrit par le simple mouvement d'un élément, on parle de scrambling, terme qui évoque du mouvement dans tous les sens (cf. angl. scrambled eggs 'œufs brouillés').

8.8 Connexion

La connexion est pour nous la première des notions syntaxiques. Cette notion est élémentaire et peut être appréhendée avec un minimum de moyens, avant des notions plus complexes telles que celles de dépendance ou de constituance.

Il y a CONNEXION (syntaxique) dès que deux éléments **se combinent** pour former un syntagme. Nous verrons que la connexion est une forme d'abstraction sur la notion de combinaison. Une connexion peut être réalisée par différentes combinaisons équivalentes.

La notion de connexion est due à Lucien Tesnière, qui à défaut d'en donner une caractérisation, lui attribue un rôle central dans le fonctionnement de la langue. Tesnière défend le fait que la connexion est un élément à part entière de la phrase. Au premier chapitre de son ouvrage posthume de 1959, Éléments de syntaxe structurale, on trouve :

« Il est indispensable que [les connexions] soient aperçues par l'esprit, sans quoi la phrase ne serait pas intelligible. Quand je dis : *Alfred parle*, je n'entends pas dire d'une part qu' « il y a un homme qui s'appelle Alfred » et d'autre part que « quelqu'un parle », mais j'entend dire tout à la fois que « Alfred fait l'action de parler » et que « celui qui parle est Alfred ».

Il résulte de ce qui précède qu'une phrase du type *Alfred parle* n'est pas composée de **deux** éléments 1° *Alfred*, 2° *parle*, mais bien de **trois** éléments, 1° *Alfred*, 2° *parle* et la connexion qui les unit et sans laquelle il n'y aurait pas de phrase. »

Avant cela, il introduit la connexion en ces termes :

« Tout mot qui fait partie d'une phrase cesse par lui-même d'être isolé comme dans le dictionnaire. Entre lui et ses voisins, l'esprit aperçoit des connexions, dont l'ensemble forme la charpente de la phrase. [...] Comprendre une phrase, c'est saisir l'ensemble des connexions qui en unissent les différents mots. »

Nous prenons totalement à notre compte cette citation de Tesnière, à un point près, le rôle prédominant donné au mot. Comme nous l'avons déjà dit, nous ne considérons pas que le mot soit l'unité minimale de la syntaxe. Reconnaître des mots, c'est déjà connecter entre eux des syntaxèmes. Et inversement, des unités plus large que le mot peuvent être combinées entre elles et les connexions peuvent donc être réalisées entre différents types d'unités.

Une remarque terminologique : Tesnière ajoute que la connexion est hiérarchisée, unissant un gouverneur à un dépendant (un *subordonné* dans ses termes). De nos jours, une relation hiérarchisée de ce type est appelée une *dépendance*. Pour notre part, nous distinguons la connexion (non hiérarchisée a priori, éventuelle-

ment hiérarchisable) de la DÉPENDANCE (une connexion hiérarchisée), contrairement à Tesnière pour qui une connexion se saurait être que hiérarchisée. Par ailleurs, Tesnière oppose la connexion à d'autres types de combinaisons comme la jonction ou la translation. Pour nous, la connexion décrit tout type de combinaison syntaxique.

8.9 Unité syntaxique

Nous avons appelé **syntagme** toute combinaison de **syntaxèmes**. Une **unité syntaxique** est soit un syntaxème, soit un syntagme (voir Chapitre 7).

Définir la connexion, c'est d'abord décider quelles portions d'un texte forment une unité syntaxique. Rappelons la définition d'une unité syntaxique.

Une unité syntaxique est : une portion de l'énoncé qui forme un **signe linguistique**

et qui **commute librement** dans l'énoncé ou qui est analogue à une telle unité.

Elle possède donc un signifiant qui est une portion de l'énoncé et un signifié qui est le sens que l'on peut attribuer à cette portion.

Il peut s'agir d'une **portion discontinue** de l'énoncé, même si une majorité d'unités syntaxiques sont des portions continues de l'énoncé. Par exemple, *le*

chien est considéré comme une sous-unité syntaxique de le petit chien, car la suppression de petit ne modifie pas l'interprétation de $le \oplus chien$. De même, le livre est sur la table est une sous-unité syntaxique de le livre de syntaxe est sur la table.

Par contre, il s'agit d'une **portion connexe** de l'énoncé du point de vue de la combinaison des signes, c'est-à-dire qu'à l'intérieur du syntagme toutes les sous-unités se sont combinées ensemble, directement ou indirectement. Si l'on reprend notre exemple *le livre de syntaxe est sur la table*, le segment *syntaxe est sur la table* n'est pas un syntagme car *syntaxe* seul ne peut se combiner avec *est sur la table* (c'est le livre qui est sur la table).

Il se peut que le signifiant d'une unité syntaxique ne soit pas une portion délimitable de l'énoncé notamment lorsqu'il y a en jeu un amalgame ($au = \dot{a} + le$) ou un syntaxème zéro (voir l'encadré ?? sur *Syntaxème zéro et troncation en fran*çais).

En l'absence de figement sémantique, une unité syntaxique prise isolément ou dans un autre contexte doit posséder le même sens que celui qu'elle a dans le contexte de l'énoncé. Prenons un exemple : Pierre a bu une bouteille de vodka entière. On peut extraire le segment vodka entière et trouver éventuellement un sens à ce segment hors contexte (la vodka entière est meilleure pour la santé). Néanmoins quel que soit le sens qu'on lui donne, il n'a pas cette contribution dans l'énoncé considéré et n'est donc pas une unité syntaxique de cet énoncé. Par contre, une bouteille entière est bien une unité syntaxique de l'énoncé. Elle commute librement avec le reste de l'énoncé et peut par exemple être remplacée par une demi-bouteille ou une grande quantité (Pierre a bu une demi-bouteille de vodka; Pierre a bu une grande quantité de vodka).

Dans la délimitation des unités syntaxiques, il est essentiel de savoir détecter quand une unité syntaxique garde le même sens après effacement ou insertion d'une autre unité. Nous verrons que c'est ni toujours facile, ni bien formalisable et des différences dans l'analyse d'un énoncé peuvent venir d'un simple désaccord sur le statut d'un segment comme unité syntaxique. Les exemples suivants illustrent cette difficulté.

- (7) Jean-Claude déteste le coca éventé.
- (8) Jean-Claude n'aime pas le café.

Nous considérons que le segment Jean-Claude déteste le coca est bien une unité syntaxique de l'énoncé (7), car il possède la même interprétation sans ou avec le

mot éventé, bien qu'il se peut que Jean-Claude aime le coca avec bulles. Le problème vient du fait que le coca éventé dénote une partie seulement de ce que dénote le coca. Pareillement, on considère que la négation intervient dans l'énoncé (8) sans que le reste de l'énoncé ne change d'interprétation. On veut dire par là que le sens de Jean-Claude aime le café reste inchangé quand cette proposition est niée : la phrase Jean-Claude n'aime pas le café est synonyme de il est faux que Jean-Claude aime le café et le sens de l'énoncé (8) ou de sa paraphrase est bien obtenu par la composition du sens de Jean-Claude aime le café et de la négation (ne...pas ou il est faux que).

Nous allons continuer à préciser la définition des unités syntaxiques et donner quelques-unes de leurs propriétés caractéristiques.

8.10 Test de commutation

La COMMUTATION avec d'autres segments est un bon critère pour voir si l'on a affaire à une unité syntaxique, puisque les unités syntaxiques, par définition, commutent librement dans leur contexte (commutation qui, rappelons-le, doit préserver le sens du contexte) (voir section 5.2 Commutation et exclusion mutuelle).

TEST DE COMMUTATION. Une unité syntaxique appartient à un paradigme d'unités syntaxiques de même distribution. Si le segment A commute avec un segment qui est clairement une unité syntaxique, alors A est potentiellement une unité syntaxique.

En particulier, le fait de pouvoir remplacer une portion de l'énoncé par un mot est un bon indice du fait qu'il s'agit d'une unité syntaxique, puisque le mot est une unité syntaxique. Un test courant, appelé TEST DE PRONOMINALISATION, est de vérifier la commutation avec un pronom : tout segment qui commute avec un pronom est une unité syntaxique. Ce test permet en plus de caractériser certains types d'unités selon les pronoms avec lesquels elles commutent. Nous y reviendrons dans la section ?? sur Les tests de constituance.

Considérons l'énoncé suivant et demandons-nous quels sont les segments de deux mots qui forment une unité syntaxique :

(9) Cet étudiant veut acheter un livre de syntaxe.

Cet énoncé contient quatre segments de deux mots qui commutent avec un mot :

Table 8.1

Cet étudiant	veut acheter	un livre	de syntaxe
II .	achète	celui	neuf

Chacun de ces segments est bien une unité syntaxique. Attention, ce critère ne suffit pas : par exemple, *acheter un* peut commuter avec *parler* et donne une phrase acceptable (*Il veut parler livre de syntaxe*, au sens de *Il aime parler politique*), mais *acheter un* n'est pas clairement une unité syntaxique de cet énoncé, car on ne peut pas l'utiliser isolément, ni l'utiliser dans un autre type de contexte.

Une unité syntaxique ne commute pas toujours avec un mot. Si l'on prend l'énoncé *Cet étudiant possède un très vieux livre*, on ne peut pas remplacer l'unité syntaxique *vieux livre* par un seul mot, car il n'existe pas de mot qui puisse se combiner à la fois avec *un*, *très* et le verbe *possède*.

Une unité syntaxique dans un contexte donné appartient à un paradigme d'unités qui peuvent commuter avec elle et qu'on appelle son paradigme de commutation. Les unités qui peuvent occuper les mêmes positions et appartiennent donc aux mêmes paradigmes de commutation forment une CLASSE DISTRIBUTIONNELLE.

Comme on le verra dans la suite, il existe un nombre restreint de grandes classes distributionnelles et le repérage de la classe distributionnelle d'une portion d'énoncé est un critère pour décider s'il s'agit d'une unité syntaxique.

Un exemple intéressant et problématique est le cas du syntagme *le petit livre*. Ici les trois fragments envisageables, *le livre*, *petit livre* et *le petit* sont tous acceptables. Mais le fragment *le petit* est-il réellement une sous-unité syntaxique de *le petit livre*? (Nous parlons de l'unité où *petit* est bien un adjectif comme dans *Je n'aime pas le grand livre*, *je prends le petit*.) La réponse est plutôt non et c'est la distribution de ce fragment qui est en jeu : *le petit* existe bien comme unité

syntaxique du français, mais cette unité n'a pas la même distribution que le fragment le petit dans le petit livre, même si elle paraît avoir à peu près le même sens. En effet, l'unité syntaxique le petit a une distribution de groupe substantival : Le petit est sur la table, Je préfère le petit. Il commute alors avec des pronoms comme il, ça et d'autres groupes substantivaux comme ce truc, ce qu'il ne peut pas faire lorsqu'il se combine avec livre dans le petit livre. Notons au passage que le fait que les adjectifs puissent former des groupes substantivaux avec un déterminant est une propriété spécifique du français, que ne possède pas par exemple l'anglais (the little ne peut pas être utilisé sans nom), sur laquelle nous reviendrons dans le Chapitre 9 et notamment l'encadré ?? sur la Co-occupation.

8.11 Unité syntaxique autonomisable

Nous avons plusieurs fois considéré une portion d'énoncé isolément pour décider s'il s'agit d'une unité syntaxique. C'est l'AUTONOMISABILITÉ ILLOCUTOIRE.

Un signe linguistique est dit illocutoirement autonomisable s'il peut être muni d'une prosodie appropriée et **former un énoncé autonome**, tel que « *Demain.* » ou « *Un chat!* ». Un énoncé autonome est un signe linguistique qui forme à lui seul un **tour de parole complet** d'un locuteur.

On considère en particulier que la réponse à une question est un énoncé autonome, même si à proprement parler celle-ci ne forme une unité sémantique complète qu'avec la question. On aura ainsi comme énoncés autonomes : (Combien sont-ils?) Deux; (Comment est-elle?) Plutôt sympa; (Quel est son métier?) Professeur.

Toutes les unités syntaxiques ne sont pas illocutoirement autonomisables, mais un grand nombre le sont et notamment les propositions, groupes substantivaux, groupes prépositionnels, groupes adjectivaux, etc.

Test d'automisabilité illocutoire. Tout portion (continue ou non) d'un

énoncé qui est illocutoirement autonomisable, c'est-à-dire qui peut **former un énoncé autonome** sans modification de sens, est une unité syntaxique.

Il s'agit d'un critère suffisant pour être une unité syntaxique, mais pas nécessaire. Ce que de nombreux linguistes appelle le groupe verbal (voir section ??) forme une unité syntaxique, mais n'est en général pas illocutoirement autonomisable. Par exemple, si l'on reprend l'exemple de la section 8.10, cet étudiant, un livre ou de syntaxe sont illocutoirement autonomisables (C'est un livre de quoi? De syntaxe), mais pas veut acheter ou veut acheter un livre de syntaxe ou veut, que nous classons pourtant intuitivement parmi les unités syntaxiques.

Pour inclure ces dernières unités, on peut relâcher le critère d'autonomisabilité et considérer non seulement les unités qui peuvent former un tour de parole, mais aussi celles qui peuvent former une **unité prosodique** dans un tour de parole.

TEST D'AUTOMISABILITÉ PROSODIQUE. Tout segment d'un énoncé qui est prosodiquement autonomisable, c'est-à-dire qui peut **former une unité prosodique** sans que cela modifie le sens de l'énoncé, est une unité syntaxique.

Ce test repose sur la définition préalable des unités prosodiques, ce qui n'est pas sans poser problème. Nous nous contenterons d'une notion intuitive de cohésion prosodique, basée sur la perception d'une coupure mélodique marquée ou non par une pause. Les frontières prosodiques sont essentiellement marquées en français par l'allongement de la dernière syllabe (la dernière voyelle est tenue plus longtemps qu'ailleurs), des variations importantes de la fréquence fondamentale et pour les frontières majeures des pauses, c'est-à-dire une interruption du son de l'ordre du dixième de seconde au moins.

Le critère d'autonomisabilité prosodique ne demande pas que le segment considéré forme une unité prosodique effective de l'énoncé, mais que l'énoncé ou une partie puisse être prononcé avec une prosodie telle que le segment forme un groupe prosodique (sans changer le contenu informationnel; voir la section 2.3 *Partir d'un sens*). On peut donc se baser sur une prosodie suffisamment « propre »

pour que les groupes prosodiques soient clairement marqués par des coupures appuyées et que leur définition ne pose pas de problèmes. On voit que ce critère résout le problème des formes verbales finies et des différentes unités qui commencent par une forme verbale finie. En effet, on peut facilement marquer une frontière prosodique entre cet étudiant et veut et grouper veut avec acheter ou acheter un livre :

(10) Cet étudiant | veut acheter | un livre de syntaxe.

Le critère d'autonomisabilité prosodique repose sur l'hypothèse que les groupes prosodiques sont nécessairement des unités syntaxiques. Ceci est en grande partie vrai, mais quelques problèmes arrivent. En particulier, les clitiques, qui sont des mots non accentués qui ne peuvent former à eux seuls une unité prosodique, peuvent s'accoler à des mots auxquels ils ne sont pas syntaxiquement liés (voir Chapitre ??). Le meilleur exemple en français est la conjonction de subordination que dans un exemple comme je pense | qu'hier | il n'est pas venu. En effet, qu'hier forme un groupe prosodique, mais pas une unité syntaxique : comme nous le verrons plus tard, la conjonction de subordination que se combine avec la construction verbale pour former une proposition subordonnée et pas avec le complément de temps hier. En anglais, le marqueur du génitif saxon 's peut s'accoler à un groupe substantival assez complexe et il formera avec le denier mot de ce segment un groupe prosodique même si ce n'est pas avec lui seul qu'il se combine syntaxiquement : the king | of England's | grandmother 'la grand-mère du roi d'Angleterre'. En effet, le syntaxème 's se combine syntaxiquement avec the king qui est le nom principal du groupe substantival the king of England. En dehors de tels cas, le critère semble valable.

On aura noté que l'autonomisabilité illocutoire et l'autonomisabilité prosodique sont de bons indices pour identifier les unités syntaxiques, mais ce ne sont en aucun cas des propriétés nécessaires. En particulier, les unités syntaxiques en deçà du mot, et notamment les syntaxèmes flexionnels, ne vérifient pas du tout ce type de propriétés comme on le verra dans la Partie 4. Dans le Chapitre ?? consacré aux *Mots*, nous introduirons la notion d'autonomisabilité faible qui nous permettra de caractériser les mots. Dans la suite, nous travaillerons avec les unités autonomisables et incluront les mots, même si certains mots comme les articles, les prépositions ou les conjonctions ne sont autonomisables ni illocutoirement, ni prosodiquement.

8.12 Représenter les combinaisons

Nous avons défini ce que nous entendions par unité syntaxique. Il y a connexion dès que deux unités syntaxiques se combinent pour former une unité syntaxique. Une telle unité syntaxique, contenant une connexion, est un syntagme. La définition des unités syntaxiques et des connexions sera précisée à plusieurs reprises dans la suite : ici même dans la section 8.18 Tests pour la connexion, dans les chapitres suivants où nous enrichirons notre bagage conceptuel avec les notions de dépendance (Chapitre 9), de constituants syntaxiques majeurs (Chapitre 10) et de constituants topologiques (Chapitre 11), mais aussi dans les parties suivantes, où nous étudierons en détail les unités syntaxiques des différents niveaux de cohésion syntaxique (nanosyntaxe, microsyntaxe et macrosyntaxe). Ce qui nous intéresse maintenant, c'est de pouvoir représenter graphiquement les combinaisons et les connexions. Nous allons voir qu'il y a deux façons de représenter une combinaison.

Supposons que A et B soient deux unités syntaxiques qui se combinent, comme par exemple, A = cet étudiant veut acheter et B = un livre de syntaxe. Nous notons $A \oplus B$ la combinaison de A et B. Nous proposons deux façons de représenter la connexion entre A et B:

de représenter la combinaison $A \oplus B$

Figure 8.2 - Deux façons

⊕ un livre de syntaxe

Figure 8.3 – Deux représentations pour cet étudiant veut acheter

À gauche, la combinaison est représentée par une **bulle** qui entoure les deux éléments qui se combinent. À droite, la combinaison est représentée par un **trait** qui lie les deux éléments qui se combinent. C'est équivalent, il s'agit juste de conventions de représentation différentes. Néanmoins, les deux conventions sont utilisés par des théories différentes et conduisent généralement à des analyses différentes, comme nous allons le voir.

8.13 Composition et décomposition

Raisonner en termes de **composition**, c'est avoir une **approche ascendante** (angl. *bottom-up*) de la combinaison : les éléments se combinent pour former des unités de plus en plus grosses.

À l'inverse, la **décomposition** est une approche **descendante** (angl. *top-down*). On part d'une unité qu'on estime complexe, comme la phrase (à supposer qu'on sache la définir), et on décompose en morceaux de plus en plus petits. Autrement dit, si $U = A \oplus B$, on peut soit dire que A et B se combinent pour former B (vision ascendante), soit dire que B0 se décompose en A1 et B2 (vision descendante).

Raisonner en termes de composition n'est pas complètement équivalent à raisonner en termes de décomposition. En effet, dire que A et B se combinent, revient à postuler une connexion entre A et B, mais n'oblige pas vraiment à considérer le tout que forment A et B ensemble. La décomposition, elle, oblige par essence à considérer le tout, dont on part pour trouver les parties. La distinction apparaît plus clairement lorsqu'on considère trois éléments A, B et C, où B se combine à la fois avec A et C (comme dans *Marie regarde Pierre* avec A = Marie, B = regarde, C = Pierre). En termes de composition, on peut représenter la chose de l'une des deux façons suivantes :

Figure 8.4 – Deux façons de représenter le même ensemble de connexions

La deuxième représentation, avec des bulles qui s'enchevêtrent, n'est pas usuelle et nous l'éviterons également.

Si l'on raisonne en termes de décomposition, on partira du tout U que forment A, B et C (U = ABC). On aura alors trois façons de décomposer U : 1) une décomposition ternaire U = A \oplus B \oplus C; 2) une décomposition binaire U = A \oplus BC; et 3) une décomposition binaire U = AB \oplus C.

de décomposer récursivement $A \oplus B \oplus C$

Figure 8.5 – Les trois seules façons

Aucune des trois décompositions n'est équivalente à l'ensemble de connexions que nous voulons considérer. Décomposer en trois ne permet pas d'indiquer que c'est bien B qui se combine avec A et C, mais que A et C ne se combinent pas directement. Décomposer en deux, par exemple en $A \oplus BC$, ne permet pas d'indiquer que c'est avant tout B qui se combine avec A. Décomposer en deux suppose en fait d'ordonner les deux connexions, $A \oplus B$ et $B \oplus C$: par exemple, on combine d'abord B avec C, puis le tout avec A. C'est ce qu'on appelle la STRATIFICATION (nous y reviendrons dans la section 10.15 sur l'*Arbre de constituants binaire*). Si la décomposition binaire a au moins l'avantage de montrer qu'il y a bien deux combinaisons binaires et pas une combinaison ternaire, elle ne permet pas de rendre compte que ces deux combinaisons sont indépendantes et qu'il n'y a pas emboî-

tement de l'une dans l'autre. Raisonner en termes de décomposition oblige à stratifier, c'est-à-dire à ordonner les connexions. C'est une information supplémentaire dont l'utilité est discutable. À l'inverse, on perd de l'information, puisque si on a toujours deux connexions, les combinaisons qui leur correspondent sont moins fines.

L'analyse descendante de la structure syntaxique a engendré l'analyse en constituants immédiats (voir section ?? éponyme). Dans cet ouvrage, nous privilégions une analyse ascendante, tout en exploitant les avantages d'une analyse descendante (voir la section ?? *Combiner les méthodes ascendante et descendante*).

8.14 La connexion et ses instances

Jusqu'à présent, nous avons considéré qu'à toute combinaison de deux unités syntaxiques correspondait une connexion, et inversement. Imaginons que l'on veuille représenter toutes les combinaisons possibles entre toutes les unités syntaxiques d'un énoncé comme *Cet étudiant veut acheter un livre de syntaxe*. C'est possible, mais pas nécessaire. Une telle représentation serait très redondante, parce que certaines **combinaisons** sont fondamentalement **équivalentes** et qu'en donner une permet d'avoir les autres. Expliquons-nous.

On peut décomposer notre phrase en cet étudiant veut acheter \oplus un livre de syntaxe, ce qui revient à dire que A = cet étudiant veut acheter et B = un livre de syntaxe sont deux unités syntaxiques qui se combinent pour donner la phrase complète. Mais acheter un livre de syntaxe est aussi une unité syntaxique de la phrase que l'on peut décomposer en acheter \oplus un livre de syntaxe. De même, cet étudiant veut acheter un livre est une unité syntaxique de la phrase décomposable en cet étudiant veut acheter \oplus un livre. En un sens, ces combinaisons correspondent à une seule et même connexion qui est celle du verbe acheter avec son complément d'objet direct.

Figure 8.6 - Combinaisons équivalentes

Nous dirons donc que ces combinaisons sont équivalentes et qu'elles sont des instances d'une seule et même connexion. Une connexion repré-

sente donc un ensemble de combinaisons équivalentes. Il s'agit d'une « abstraction » de la notion de combinaison.

D'aucun pourrait considérer qu'il aurait été beaucoup plus simple de définir la connexion comme une relation entre les mots et donc de ne regarder que les combinaisons de deux mots. C'est ce qui est habituellement fait dans les présentations de la syntaxe de dépendance. Deux raisons nous ont amené à procéder autrement. La première raison est que nous ne souhaitons pas présupposer que les connexions ont lieu entre les mots, parce que cela nous oblige à une définition préalable des mots, ce qui n'a rien de trivial, et que, par ailleurs, nous pensons que certaines connexions ne se réalisent pas entre les mots. La deuxième raison est que nous pensons que la connexion est fondamentalement une classe d'équivalence de combinaisons (voir définition dans l'encadré ??) et que voir la connexion comme opérant seulement au niveau des mots est une vue parcellaire. On peut très bien envisager la connexion à d'autres niveaux de granularité, au niveau plus fin des syntaxèmes, comme à un niveau plus grossier (par exemple un livre ⊕ de syntaxe), voire comme les analyses en constituants immédiats, au niveau des constituants (voir Chapitre 10). Un arbre de dépendance et un arbre de constituants peuvent très bien représenter exactement les mêmes connexions, mais ils le font avec des instances différentes : un arbre de dépendance représente les connexions par des combinaisons entre mots, tandis qu'un arbre de constituants ordonne les connexions et les représentent par des enchâssements de constituants.

Notre définition permet d'appréhender plus clairement une question que l'on se pose généralement lorsqu'on aborde la syntaxe de dépendance : « Les dépendances syntaxiques ont-elles réellement lieu entre les mots? ». Nous pensons qu'on ne peut pas et qu'on ne doit pas répondre à cette question. Cette question n'a pas lieu d'être et c'est ce que montre notre définition de la connexion et celle de la dépendance qui en découle. Une connexion ne lie pas plus deux mots qu'elle ne lie toutes les unités plus larges qui incluent ces deux mots, tout comme elle peut lier des unités qui se trouvent en deçà du mot. Tout au plus peut-on se demander si les instances minimales des connexions lient des mots, mais la réponse à cette question varie selon les connexions et selon les critères définitionnels des unités syntaxiques choisis. Autrement dit, le fait que l'instance minimale d'une connexion lie deux mots n'est pas vraiment une propriété de la connexion en soi, mais bien davantage une conséquence des critères retenus pour définir les unités

syntaxiques en général, lesquels critères sont ajustables et en partie indépendants des critères qui permettent de postuler une connexion donnée.

8.15 Définition formelle de la connexion



Nous avons défini ci-dessus une connexion comme un ensemble de combinaisons équivalentes. Cette définition peut être précisée. Commençons par donner une définition un peu plus formelle de l'équivalence entre deux combinaisons.

Une combinaison $A \oplus B$ est plus fine qu'une combinaison $A' \oplus B'$ si A est inclus dans A' et B est inclus dans B'. Par exemple, la combinaison veut acheter \oplus un livre est plus fine que la combinaison veut acheter \oplus un livre de syntaxe.

Deux combinaisons sont ÉQUIVALENTES s'il existe une combinaison plus fine que les deux à la fois. Par exemple, acheter ⊕ un livre de syntaxe et cet étudiant veut acheter \oplus un livre sont équivalentes car la combinaison acheter \oplus un livre est plus fine que les deux.

Dès qu'on a une RELATION D'ÉQUIVALENCE sur un ensemble, on peut considérer les objets à l'équivalence près. (Nous donnons une définition plus formelle d'une relation d'équivalence dans l'encadré ?? sur Dépendance, dominance et transitivité.) Une relation d'équivalence partitionne l'ensemble des objets en classes D'ÉQUIVALENCE. Donnons un exemple classique, celui des nombres rationnels. Les nombres rationnels sont les nombres qui peuvent s'écrire à l'aide d'une fraction comme $\frac{3}{4}$ ou $\frac{257}{23}$. Les fractions $\frac{3}{4}$ ou $\frac{75}{100}$ sont deux fractions différentes qui représentent le même nombre rationnel. Elles sont donc équivalentes. Un nombre rationnel est une classe d'équivalence de fractions et il peut s'écrire d'une infinité de façons. De la même façon, les combinaisons sont les instances des connexions et une connexion peut être représentée par des combinaisons mettant en jeu des unités plus ou moins grandes.

On peut maintenant définir la connexion à partir de la relation d'équivalence que nous venons de définir sur les combinaisons. De même qu'un nombre rationnel est une classe d'équivalence de fractions, une connexion est une classe d'équivalence de combinaisons. Cette courte définition cache, on l'a compris, une opération d'abstraction complexe.

8.16 Tests pour la connexion

Les notions de connexion et d'unité syntaxique sont intimement liées. Dire qu'il y a une connexion entre les unités syntaxiques A et B revient à dire que AB ou BA forme une unité syntaxique. Les tests pour la connexion sont donc des tests pour vérifier qu'une portion d'un énoncé forme une unité syntaxique.

Nous avons donné trois tests pour décider si un segment de texte est une unité syntaxique :

- le test de commutation (section 8.10)
- le test d'autonomisabilité illocutoire (section 8.11)
- le test d'autonomisabilité prosodique (section 8.11)

On verra dans le Chapitre 10 des tests suffisants pour caractériser une unité syntaxique, mais qui ne s'appliquent qu'à un sous-groupe d'unités syntaxiques, comme les constituants majeurs (section ?? Tests de constituance).

L'ORDRE DES MOTS (que nous préférons appeler la TOPOLOGIE, notamment car cela ne concerne pas que les mots; voir Chapitre 11) est une des propriétés qui caractérise les connexions. Les principes généraux sont que l'ordre des unités syntaxiques dépend fortement des connexions : les unités syntaxiques qui sont connectées tendent à se placer les unes par rapport aux autres et les unes à côté des autres. Ceci nous donne deux tests.

TEST DE DÉPLACEMENT. Plus les placements des unités syntaxiques A et de B sont dépendants l'un de l'autre, plus il y a des chances que A et B soit connectés.

TEST D'INSERTION. Si on ne peut pas insérer entre A et B d'unité qui ne soit pas connectée à A et B, alors A et B sont probablement connectés.

Les énoncés des tests que nous avons donnés sont très généraux. Ils vont trouver des déclinaisons particulières selon les langues étudiées et les constructions concernées.

En français, l'usage le plus courant du test de déplacement concerne les dépendants du nom, qui, à l'exception de certains adjectifs, se placent tous à droite du nom dont ils dépendent. Nous allons donner deux illustrations. La première concerne les propositions :

- (11) Pierre a acheté une glace à la boulangerie.
- (12) Pierre a acheté une glace à la fraise.

Dans les deux exemples, on a un groupe prépositionnel B introduit par \grave{a} dont on cherche à savoir s'il est connecté à A = une glace. La permutation de A et B offre immédiatement une réponse :

- (13) Pierre a acheté à la boulangerie une glace.
- (14) *Pierre a acheté à la fraise une glace.

La permutation est possible dans le premier cas, donc à la boulangerie ne dépend pas de une glace. Par contre, elle est impossible dans le deuxième cas, ce qui indique que à la fraise dépend bien de une glace. (Nous utilisons ici une autre propriété, spécifique au français, qui est que l'ordre des compléments à droite du verbe est relativement libre. Le même test serait plus difficile à appliquer dans une langue comme l'anglais où l'objet ne peut pas être facilement séparé du verbe.)

Figure 8.7 – Deux structures syntaxiques différentes

Le test de déplacement est particulièrement utile pour déterminer les dépendances au sein du groupe substantival (où les tests de pronominalisation ou de clivage ne peuvent pas être utilisés; voir Chapitre 10 pour la présentation de ces tests). Considérons l'exemple suivant (extrait de l'article *Seine* de la wikipédia) :

(15) La faible déclivité de la vallée de la Seine en Ile-de-France a causé la formation de multiples et profonds méandres.

On a ici trois segments prépositionnels, *de la vallée*, *de la Seine* et *en Ile-de-France*, dont on doit déterminer le gouverneur. Le test de déplacement donne :

- (16) *la faible déclivité de la Seine de la vallée en Ile-de-France
- (17) la faible déclivité en Ile-de-France de la vallée de la Seine

La non-déplaçabilité du segment *de la Seine* indique qu'il est bien connecté au segment *de la vallée*. À l'inverse la déplaçabilité de *en Ile-de-France* indique qu'il n'en dépend pas. On a donc les connections suivantes :

Le test d'insertion peut être appliqué à notre premier exemple. Prenons un circonstanciel comme l'autre jour, qui dépend du verbe, mais ne se connecte pas avec les unités A et B que nous considérons. On a :



Figure 8.8 - Rattachement des segments prépositionnels

- (18)Pierre a acheté une glace l'autre jour à la boulangerie.
- ?*Pierre a acheté une glace l'autre jour à la fraise. (19)

(Des énoncés tels que le deuxième sont tout à fait possibles avec une prosodie appropriée, où le segment à la fraise est détaché. Mais ils ne pourraient être produits à l'écrit sans une ponctuation marquant le détachement.) Le fait que l'autre jour peut être inséré entre A et B dans le premier cas et pas dans le deuxième indique que A et B sont connectés dans le premier cas et pas dans le deuxième.

De même, dans notre deuxième exemple, on peut vérifier que en Ile-de-France ne peut pas être déplacé entre A = de la vallée et B = de la Seine, et donc que A et B son connectés:

(20)*la faible déclivité de la vallée en Ile-de-France de la Seine

Le test d'insertion est un cas particulier de la propriété de projectivité que nous présentons dans la section 11.15 Projectivité et dépendance projective. Cela revient à dire que certains liens de connexion ne peuvent pas se couper et que la configuration ci-dessous n'est pas possible.

Figure 8.9 – Configuration rejetée par le test d'insertion

Les deux tests que nous venons d'énoncer, le test de déplacement et le test d'insertion, sont des tests qui s'appliquent seulement à certaines connexions et qui dépendent de propriétés générales de la langue et de propriétés particulières de certaines constructions. Par exemple, si le test d'insertion s'applique bien à la connexion entre un nom et un complément de nom en français, il n'en sera pas de même dans des langues qui acceptent plus facilement des configurations non projectives ou pour d'autres constructions du français (voir la section ?? Non-projectivité en français).

8.17 Connexion binaire?



Notre définition de la connexion présuppose que toute **connexion** est **binaire**, c'est-à-dire que les unités syntaxiques se combinent deux par deux. Ceci revient à supposer qu'on peut décomposer toute unité en deux morceaux. Existe-t-il des unités qui ne peuvent être décomposées qu'en trois morceaux?

Considérons l'énoncé il a invité plein de gens. En première analyse, plein de gens ne peut pas être décomposé en deux morceaux. Il contient trois morceaux qui forment des unités syntaxiques — plein, de et gens —, mais aucune combinaison deux à deux de ces éléments ne forme un segment autonomisable : *plein de, *de gens, *plein gens. Si tel est le cas, il faut envisager une connexion ternaire, c'est-à-dire une connexion à trois sommets.

La syntaxe de dépendance repose traditionnellement sur le présupposé que toutes les combinaisons sont binaires. Cette hypothèse mérite d'être discutée. On peut évidemment toujours réduire (artificiellement) une connexion ternaire à deux connexions binaires. Par exemple dans notre exemple, considérer que de et gens se combinent ensemble avant de se combiner avec plein ou au contraire considérer que plein et de se combinent ensemble avant de se combiner avec gens. Les deux analyses ont été proposées; la première revient à faire de de gens un groupe prépositionnel (et se justifie par la pronominalisation en en (il en a invité plein) et par des coordinations comme plein de chats et de chiens), tandis que la deuxième fait de plein de un déterminant complexe (et se justifie par la commutation de plein de avec des déterminants simples comme des ou plusieurs) (voir l'encadré ?? sur les « Déterminants complexes »).

Du point de vue mathématique, les connexions ternaires ne posent pas de problème en soi, mais elles nécessitent d'introduire des structures légèrement différentes des arbres et graphes habituellement utilisés en modélisation des langues. (Un graphe, au sens usuel, ne contient que des connexions binaires.) Considérer que *plein de gens* est une connexion ternaire revient à considérer que *plein* et *gens* se combinent directement et qu'ils ne peuvent se combiner qu'en présence du marqueur *de*, ce qui est finalement assez facile à concevoir. On peut par exemple représenter cela par un lien entre *plein* et *gens* surmonté par *de*:

Dans une telle représentation, le trait de connexion lie bien trois éléments et représente donc une connexion ternaire. Néanmoins, la position particulière donnée à *de* revient à lui donner un statut particulier, qui peut être justifiée par le fait que *de* ne peut pas entretenir d'autres liens et joue uniquement un rôle de marqueur dans la combinaison de deux lexèmes.

Pour notre part, nous travaillerons essentiellement avec des connexions binaires, mais nous garderons à l'esprit des représentations telle que la précédente, qui ont l'avantage de ne pas obliger à privilégier un des deux découpages binaires possibles pour *plein de gens*. De telles représentations restent assez simples et intuitives et elles ont été utilisées par certains auteurs. Nous discuterons d'une représentation de ce type utilisée pour la coordination par Tesnière au Chapitre ??.

8.18 Structures de connexion, granularité et critères

Une STRUCTURE DE CONNEXION est une représentation de la structure de l'énoncé contenant une instance de chaque connexion. Il existe plusieurs structures de connexion possibles selon la granularité de l'analyse et les critères pris en compte.

La Granularité de l'analyse dépend de la taille des unités minimales considérées. Pour une granularité donnée, la structure de connexion contient l'Instance minimale de chaque connexion, c'est-à-dire l'instance reliant les plus petites unités considérées pour cette granularité. L'analyse syntaxique la plus fine prendra comme unités minimales les syntaxèmes. Il est courant que les analyses syntaxiques prennent comme unités minimales les mots. Une analyse au niveau de la structure discursive d'un texte peut prendre comme unités minimales les unités illocutoires (c'est-à-dire plus ou moins les phrases, voir Partie 6). Une représentation syntaxique devra en tout cas contenir une instance de chaque connexion considérée.

La nature des critères pris en compte amène également à considérer différentes structures de connexion. Prenons l'exemple de l'énoncé Marie regardait Pierre. Cet énoncé contient, en plus des mots, les deux segments autonomisables *Marie regardait* et *regardait Pierre* et on a donc les combinaisons *Marie* \oplus *regar*dait et regardait

Pierre. Si l'on s'en tient aux unités syntaxiques autonomisables, on ne peut pas raffiner davantage ces connexions, mais on peut essayer en appliquant d'autres critères. La forme verbale regardait peut être décomposée en regard- ⊕ -ait (c'est-à-dire un lexème et sa flexion). Les deux signes sont indissociables l'un de l'autre, donc on ne peut pas espérer former des unités autonomisables avec seulement l'un ou l'autre. En revanche, on peut voir que la commutation de -ait avec d'autres flexions verbales n'a aucune incidence sur la possibilité de combiner la forme verbale avec l'objet Pierre, alors que la commutation de regard- avec d'autres lexèmes verbaux (parl- ou rigol-) rend cette combinaison impossible. On en déduit que c'est regard- qui se combine avec Pierre, car c'est lui qui contrôle la possibilité de se combiner ou pas avec Pierre. Inversement, la commutation de regard- avec d'autres lexèmes verbaux ne change rien à la combinaison avec le sujet *Marie*, alors que la combinaison de *-ait* avec d'autres flexions (l'infinitif ou l'impératif) bloque la combinaison avec le sujet. On en déduit que c'est -ait qui se combine avec Marie. Le critère que nous venons d'utiliser pour raffiner la connexion s'appelle le critère distributionnel et sera discuté aux sections 3.3.11 à 3.3.16.

Nous donnons dans la figure suivante deux instanciations des trois connexions entre les quatre unités considérées (*Marie, regard-, -ait* et *Pierre*). Dans la struc-

ture de gauche, nous utilisons le seul critère d'autonomisabilité des unités pour déterminer les connexions. Dans la structure de droite, nous ajoutons le critère de distributionnalité. Les deux analyses ont la même granularité, mais ne reposent pas sur les mêmes critères. La structure de droite affine les connexions de la structure de gauche en proposant des instances plus fines des connexions (voir l'encadré ?? sur la *Définition formelle de la connexion* pour la relation de finesse entre combinaisons).

Marie regardait Pierre

Structure de connexion Structure de connexion avec critère d'autonomisabilité avec critère distributionnel

Figure 8.10

Dans la suite de ce chapitre, nous travaillerons avec les unités syntaxiques autonomisables et donc avec le grain des mots, qui sont les plus petites unités syntaxiques faiblement autonomisables (voir Chapitre ??).

8.19 Structure de connexion et unités

Nous avons vu deux façons de définir une structure de connexion dans la section précédente. Dans la première, nous nous sommes entièrement basés sur les unités que l'on pouvait considérer dans la phrase Marie regardait Pierre, à savoir : Marie, Pierre, regar-, -ait, regardait, Marie regardait, regardait Pierre et la phrase entière. Dès qu'une unité est la combinaison de deux sous-unités, nous postulons une combinaison : $regardait = regard - \oplus -ait$, $Marie regardait = Marie \oplus regardait$ et regardait Pierre = regardait \oplus Pierre, mais aussi Marie regardait Pierre = Marie ⊕ regardait Pierre et Marie regardait Pierre = Marie regardait ⊕ Pierre. Certaines combinaisons sont équivalentes et correspondent à la même connexion : Marie ⊕ regardait et Marie ⊕ regardait Pierre d'une part, et regardait ⊕ Pierre et Marie regardait

Pierre d'autre part. Pour construire la structure de connexion, nous prenons l'instance minimale de chaque connexion, ce qui nous donne la structure de gauche de la section précédente. Nous avons pu construire cette structure de connexion à partir de l'ensemble des unités et de lui seul. Inversement, la structure de connexion nous permet de retrouver l'ensemble des unités. Pour cela, il faut imaginer qu'on a des ciseaux et qu'on peut couper des connexions pour ne garder que des bouts de la structure. Chaque bout obtenu est une unité. Par exemple, on peut couper le lien entre Marie et regardait et il nous reste regardait Pierre. On peut tout couper et obtenir les unités minimales, y compris



regard- et -ait. En revanche, on ne peut obtenir de combinaison de Marie avec seulement regard- ou -ait, car Marie est connecté à tout regardait. Dans le cas où la structure de connexion est définie en se basant uniquement sur des critères définissant les unités syntaxiques, nous avons une dualité parfaite entre la structure de connexion et l'ensemble des unités. (La notion de dualité est présentée dans l'encadré ?? éponyme.)

Dans la figure de droite de la section précédente, nous avons utilisé un critère supplémentaire, le critère distributionnel, pour affiner la structure et postuler les combinaisons $Marie \oplus -ait$ et $regard - \oplus Pierre$. Ces combinaisons ne sont plus basées sur des unités : nous ne considérons pas que Marie ⊕ -ait ou regard- ⊕ Pierre forment des unités. Il n'y a plus de dualité entre la structure de connexion et l'ensemble des unités syntaxiques considérées.

8.20 Les limites de la dualité



Timothy Osborne propose d'appeler CATENA ('chaîne' en latin) toute portion connexe d'un arbre de dépendance. (Le terme est un peu trompeur, puisque les catenae peuvent brancher et ne pas être stricto sensu des chaînes.) Les catenae sont toutes les unités obtenues lorsqu'on découpe l'arbre de dépendance en morceaux en coupant des dépendances. Les catenae comprennent la quasi-totalité des unités qui jouent un rôle dans la langue : non seulement tous les constituants considérés par les grammaires de constituants, y compris les projections intermédiaires (voir la section 10.15 Arbre de constituants binaire), mais aussi les formes verbales complexes (a mangé), les collocations (avoir besoin), les locutions (sur le point (de Vinf)) ou encore les nucléus verbaux dont nous montrerons le rôle dans l'extraction (le livre que tu veux que je lise) et d'autres phénomènes aux chapitres 5.3 et 5.4. Ces auteurs tirent argument du fait que ces unités peuvent ainsi être facilement déduites d'un arbre de dépendance pour justifier le fait qu'un arbre de dépendance encode la structure syntaxique de manière plus satisfaisante qu'un arbre de constituants (voir la section 8.25 sur l'Analyse en constituants immédiats).

Nous pensons néanmoins qu'il y a un problème méthodologique à vouloir partir d'un arbre de dépendance pour en déduire les unités de la langue. D'une part, la définition d'un arbre de dépendance est basée sur une structure de connexion qui est une façon de représenter de manière structurée et condensée l'ensemble des unités syntaxiques et il semble difficile de définir cette structure sans avoir défini au préalable ce qu'est une unité syntaxique; ce n'est donc pas les unités qui se déduisent de l'arbre de dépendance, c'est l'arbre de dépendance qui se déduit de l'ensemble des unités considérées. D'autre part, la définition d'un arbre de

dépendance est **plus complexe** que l'identification des unités syntaxiques et fait appel à de nombreux critères (les critères utilisés pour l'identification des unités syntaxiques, mais aussi le critère distributionnel, déjà évoqué à la section 8.18, et d'autres encore que nous introduirons au Chapitre 9). Ces critères permettent en particulier de mettre en évidence une hiérarchie qui n'est pas nécessaire pour définir les unités syntaxiques. Par ailleurs, en raison de la prise en compte de critères supplémentaires, les catenae, c'est-à-dire les portions connexes de l'arbre de dépendance, ne sont pas tous des unités syntaxiques et il n'y a donc plus de dualité parfaite entre l'ensemble des unités syntaxique et la structure de dépendance. Certaines des propriétés utilisées pour définir l'arbre de dépendance (comme le critère distributionnel) ne sont pas aisément transférables en des propriétés des unités et certains des catenae de l'arbre de dépendance n'ont pas d'intérêt théorique a priori. Par exemple, pour une phrase comme *Le chien dort*, il faudra connecter *dort* à *chien* ou à *le* pour obtenir un arbre de dépendance, mais ni *le dort* ni *chien dort* n'est une unité intéressante.

8.21 Arbre de fragmentation

Il n'est pas nécessaire d'identifier toutes les unités syntaxiques pour obtenir la structure de connexion. Une des façons d'obtenir les unités syntaxiques et la structure de connexion qui lui correspond est de procéder par décompositions successives d'un énoncé. On commence par chercher des décompositions possibles de l'énoncé en deux unités syntaxiques, puis l'on choisit une des décompositions. Pour chacune des deux unités obtenues, on recommence : on cherche des décompositions possibles en deux unités et on en choisit une. Et ainsi de suite, jusqu'à arriver aux unités minimales (par exemple les mots ou les syntaxèmes). Une série successive de décompositions est appelée une FRAGMENTATION de l'énoncé et l'ensemble des unités obtenues dans une fragmentation donnée en sont les FRAGMENTS.

Donnons des exemples de fragmentations de l'énoncé *Cet étudiant veut acheter un livre de syntaxe.* Nous limitons nos fragmentations aux unités autonomisables.

```
((\text{cet} \oplus \text{\'etudiant}) \oplus (\text{veut} \oplus \text{acheter})) \oplus ((\text{un} \oplus \text{livre}) \oplus (\text{de} \oplus \text{syntaxe}))
(\text{cet} \oplus \text{\'etudiant}) \oplus (\text{veut} \oplus (\text{acheter} \oplus (\text{un} \oplus (\text{livre} \oplus (\text{de} \oplus \text{syntaxe})))))
((\text{cet} \oplus \text{\'etudiant}) \oplus \text{veut}) \oplus ((\text{acheter} \oplus (\text{un} \oplus \text{livre})) \oplus (\text{de} \oplus \text{syntaxe}))
```

Figure 8.11 – Des exemples de fragmentation

On peut représenter une fragmentation par une structure arborescente, appe-

lée un Arbre de Fragmentation : chaque nœud de l'arbre représente un fragment, les feuilles de l'arbre étant les fragments minimaux considérés (les mots dans notre exemple). Les branches de l'arbre représente la relation entre un fragment et un des deux sous-fragments obtenu lors de la décomposition. Nous donnons l'arbre de fragmentation de la première de nos trois fragmentations.

L'arbre pour la fragmentation

 $((\text{cet} \oplus \text{\'etudiant}) \oplus (\text{veut} \oplus \text{acheter})) \oplus ((\text{un} \oplus \text{livre}) \oplus (\text{de} \oplus \text{syntaxe}))$

Figure 8.12

8.22 De la fragmentation à la connexion

On peut définir la structure de connexion d'un énoncé à partir de toutes les fragmentations possibles de cet énoncé. Il suffit de prendre toutes les combinaisons obtenues, puis de sélectionner parmi les ensembles de combinaisons équivalentes l'instance minimale (voir la section 8.18 Structures de connexion, granularité et critères). Si cette méthode est la plus efficace du point de vue théorique, ce n'est pas celle que l'on utilise en pratique. Il est inutile de considérer toutes les fragmentations, toutes les unités et toutes les combinaisons qu'induit un énoncé que l'on souhaite analyser. Il suffit en fait de calculer n'importe quelle fragmentation de l'énoncé, puis d'obtenir la structure de connexion par des raffinements successifs des combinaisons de cette fragmentation. Il suffit, pour chaque combinaison $A \oplus B$ et pour une décomposition de A en $X \oplus Y$, de regarder si l'une des combinaisons $X \oplus B$ ou $Y \oplus B$ peut raffiner la combinaison $A \oplus B$. On va ainsi raffiner chaque connexion jusqu'à obtenir une combinaison minimale qui ne peut plus être raffinée.

Prenons l'exemple de la section précédente et l'arbre de fragmentation que nous avons dessiné. On veut raffiner la combinaison $A \oplus B$ avec A = cet étudiant et B = veut acheter. A se décompose en $X \oplus Y$ avec X = cet et Y = étudiant. Aucune des deux combinaisons cet veut acheter et étudiant veut acheter ne forme une unité syntaxique et on ne peut donc raffiner ainsi. À l'inverse, B se décompose en $veut \oplus acheter$ et cet étudiant veut est une unité syntaxique. On peut donc raffiner la combinaison $A \oplus B$ en $A \oplus veut$. En continuant ainsi, on obtient la structure de connexion suivante.

Figure 8.13 – Structure de connexion

8.23 Graphe à bulles et polygraphe 🕏

Pour représenter la structure de connexion, nous avons utilisé une structure que nous proposons d'appeler un GRAPHE À BULLES. Rappelons qu'un graphe est un ensemble d'éléments, appelés nœuds, reliés deux à deux par des liens, appelés arêtes (voir l'encadré ?? sur *Graphe et arbre*). Dans un graphe à bulles, les sommets d'une arête peuvent être des bulles contenant plusieurs nœuds. Une arête qui relie des nœuds simples est dite ÉLÉMENTAIRE et une arête dont un des sommets au moins est une bulle est NON ÉLÉMENTAIRE.

Les structures de connexion ne sont pas n'importe quel graphe à bulles. D'une part, les bulles ne se chevauchent pas, c'est-dire que deux bulles sont ou bien disjointes ou bien incluses l'une dans l'autre. D'autre part, la structure a une propriété de **connexité** (voir également l'encadré ??) : la structure est formée de grandes bulles qui forment un réseau connexe et l'intérieur de chaque bulle est lui-même un réseau connexe de bulles et ceci jusqu'aux bulles élémentaires que sont les nœuds.

Enfin, les structures de connexion ont généralement la propriété de binarité suivante : chaque bulle se décompose en deux bulles connectées entre elles. Il est alors possible d'utiliser un autre mode de représentation, où les bulles ne sont pas représentées : un lien dont le sommet est une bulle composée de deux sousbulles connectées entre elles sera attribué au lien entre les deux sous-bulles. On obtient une nouvelle structure, que nous appelons un POLYGRAPHE, où une arête peut avoir pour sommet un nœud ou une arête. Nous donnons ci-dessous une version polygraphique de la structure connexion de la section précédente.

8.24 Cycle de connexions

Il peut arriver qu'une unité ABC possède comme sous-unités à la fois AB, BC et AC et que les trois combinaisons $A \oplus B$, $B \oplus C$ et $A \oplus C$ soient possibles, ce qui donne un **cycle** (non orienté) de connexions. (On verra, lorsqu'on considérera la dépendance et que les connexions sont orientées, qu'il ne peut pas s'agir d'un cycle orienté; voir la section 9.5 *Structure de dépendance et arbre de dépendance*.) Cette situation se rencontre avec les exemples suivants.

Marie commet une faute envers Pierre. la montée du nationalisme en Catalogne Je l'ai vu hier à l'école.

Dans la première phrase, on peut à la fois considérer comme acceptables les unités syntaxiques commet une faute, commet envers Pierre et une faute envers Pierre. Il s'ensuit que le rattachement du complément envers Pierre à une faute

plutôt qu'à commet est incertain, que l'un ou l'autre ne change pas le sens.

La deuxième phrase pose les mêmes problèmes. Il est assez clair que *hier* et à *l'école* sont tous deux des compléments connectés à la construction verbale (*je l'ai vu hier, je l'ai vu à l'école*), mais par ailleurs *hier à l'école* semble aussi fonctionner aussi comme unité syntaxique (*C'est hier à l'école que je l'ai vu*; *Tu l'as vu*? *Oui, hier à l'école*.)

Le dernier exemple est du même type encore. S'agit-il du *nationalisme en Catalogne* dont on évoque la montée ou bien du nationalisme dont on considère *la montée en Catalogne*. Peu importe, le sens est à peu près le même et le locuteur n'a pas vraiment les moyens de choisir une connexion plutôt que l'autre. Comme bien sûr *la montée* et *du nationalisme* sont connectés, on a, à nouveau, un cycle potentiel.

Le fait que les trois connexions sont possibles ne signifie pas que les trois connexions coexistent dans la production ou l'analyse de ces exemples. Il est probable que la prosodie produite par le locuteur privilégie une décomposition plutôt qu'une autre et donc une connexion plutôt qu'une autre. On peut penser que, une fois que le destinataire d'un tel énoncé a récupéré une structure connexe satisfaisante, il ne cherche pas à voir s'il pourrait encore la compléter pour obtenir des cycles. En conclusion, nous pensons que la structure syntaxique d'un énoncé peut être potentiellement cyclique, que la manipulation de structures cycliques ne pose pas de problèmes théoriques ou pratiques particuliers, mais nous ne pouvons pas savoir si les locuteurs manipulent ou non des structures syntaxiques cycliques lorsqu'ils produisent ou analysent un énoncé.

8.25 Analyse en constituants immédiats

La fragmentation d'un énoncé est à la base de l'ANALYSE EN CONSTITUANTS IMMÉDIATS (ACI) initiée par Leonard Bloomfield1933 et popularisée à la fin des années 1950 par la formalisation qu'en a donnée Noam Chomsky1957 (voir l'encadré ?? sur les *Grammaires de réécriture*). L'ACI fait une hypothèse supplémentaire par rapport à l'approche que nous développons : elle suppose qu'une seule fragmentation d'un énoncé est légitime et que cette unique fragmentation définit les constituants de l'énoncé. L'arbre de fragmentation correspondant est appelé un Arbre de constituants. (Cet arbre est aussi souvent appelé *arbre syntagmatique*, mais nous ne retiendrons pas ce terme qui ne correspond pas à l'usage que nous faisons, à la suite de Saussure, du terme *syntagme*. Voir l'encadré ?? À chacun son syntagme.)

Henry Gleason1955, qui est l'un des auteurs qui a défini l'arbre de constituant

avec le maximum de rigueur, est explicite sur la nécessité de choisir une fragmentation parmi toutes celles possibles. Le début de sa définition est tout à fait compatible avec la notre : « Nous pouvons, comme première hypothèse, considérer que chacun des [mots de l'énoncé considéré] a une relation énonçable avec chaque autre mot. Si l'on peut décrire ces interrelations complètement, on aura décrit la syntaxe de l'énoncé dans son entièreté. [...] On pourrait commencer par marquer ces paires de mots qui sont ressenties comme ayant les relations les plus étroites. » Puis sans justification aucune, il ajoute : « Nous allons aussi imposer la règle que chaque mot peut être marqué comme appartenant à seulement une telle paire. » Il poursuit en déclarant que la méthode pour trouver la meilleure paire parmi tous les possibles est « le problème de base de la syntaxe » (« the basic problem of syntax »), tout en admettant que sa méthode est un peu désordonnée et ne fonctionne pas complètement.

Nous ne partageons pas le point de vue de Gleason et ses contemporains. Nous avons sciemment donné une définition peu restrictive de la notion d'*unité syntaxique*. La plupart des linguistes de la seconde moitié du 20^e siècle, et notamment les courants théoriques dit générativistes, issus de l'école chomskyenne, ont défendu que seules les unités qu'ils appellent les CONSTITUANTS SYNTAXIQUES sont des unités syntaxiques légitimes. Nous pensons pour notre part que les constituants syntaxiques sont des unités syntaxiques particulières, certes remarquables, mais qu'il n'y a pas lieu de ne pas considérer les autres unités syntaxiques.

Obtenir une seule décomposition suppose d'avoir des critères beaucoup plus restrictifs que les nôtres pour la définition des unités syntaxiques. Si l'on se place dans ce cadre, on ne peut plus proposer de décomposer un livre de syntaxe à la fois en $un \oplus livre$ de syntaxe et en un livre de syntaxe. Il faut choisir entre les deux décompositions. Une telle approche suppose par exemple que l'énoncé Cet étudiant veut acheter un livre de syntaxe peut être décomposé en cet étudiant de veut acheter de syntaxe à l'exclusion de toute autre décomposition. Il faut donc proposer des critères qui fassent de ces deux fragments les deux seuls fragments acceptables (du point de vue de ces critères). Nous pensons que c'est difficile et, qui plus est, inutile. (Nous discuterons cette question en long et en large au Chapitre 10.) Au contraire, en considérant toutes les fragmentations possibles d'un énoncé, on obtient une structure plus riche et finalement plus simple.

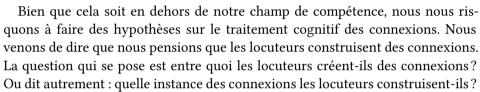
Les constituants sont des unités syntaxiques particulières, plus difficiles à caractériser que les unités syntaxiques en général. Il faut pour caractériser les constituants des critères additionnels et notamment faire intervenir la notion de



Q

tête dont il sera question au Chapitre 9. Dans le Chapitre 10, nous verrons qu'il est possible de caractériser, parmi toutes les fragmentations possibles, différentes fragmentations correspondant à un arbre de constituants et que ces fragmentations, accompagnée de l'indication des têtes, permettent de récupérer directement la structure de connexion et donc les autres fragmentations. Nous présenterons en particulier dans l'encadré ?? la Syntaxe X-barre qui est la version la plus aboutie de l'ACI.

8.26 Traitement cognitif des connexions **Q**



La syntaxe de dépendance traditionnelle répond, à la suite de Tesnière, que les connexions se font entre mots (voir la citation de Tesnière à la section 8.8). La syntaxe générative, si on suit l'analyse en constituants immédiats, répond que les locuteurs connectent les syntaxèmes pour former des constituants de plus en plus gros et que la phrase entière est obtenue par la combinaison de deux constituants que sont le sujet et le groupe verbal. Il est clair pour nous que la syntaxe de dépendance est beaucoup plus proche de la réalité cognitive. On peut néanmoins raisonnablement se demander si toutes les connexions se font entre mots. Existe-t-il des connexions qui s'instancient entre syntaxèmes de mots différents? Existe-t-il des connexions qui s'instancient entre unités plus larges que le mot?

Il est généralement admis que les locuteurs traitent les énoncés de manière IN-CRÉMENTALE, c'est-à-dire au fur et à mesure qu'ils les perçoivent. Une expérience classique est celle des GARDEN PATHS (en anglais to lead somebody up the garden path signifie 'mener quelqu'un en bateau'). Lisez les deux phrases suivantes :

- (21)L'espion reconduit à la frontière un diplomate américain.
- (22)L'espion reconduit à la frontière est un diplomate russe.

Lorsqu'on suit le mouvement des yeux d'un lecteur (on appelle cette méthode l'eye-tracking), on observe une saccade régressive sur la deuxième phrase, c'està-dire que lorsque le lecteur tombe sur la deuxième forme verbale, est, il revient sur la première, reconduit, puis revient sur est. Ceci tend à prouver que l'analyse

est globalement incrémentale, que les mots sont analysés et connectés au fur et à mesure de la lecture, mais que toute analyse qui est infirmée par la suite nécessite un retour en arrière (angl. *backtracking*) pour détricoter l'analyse en cours.

Le fait que l'analyse soit globalement incrémentale n'induit pas nécessairement que les syntaxèmes sont systématiquement traités les uns après les autres. Il est probable que les locuteurs traitent le texte par **petits blocs successifs** et que ces blocs contiennent quelques mots. Donc, d'une part, le locuteur va connecter ces blocs entre eux (voir notamment la stratégie d'analyse de la section ?? sur le *Flux de dépendances*) et d'autre part, il va les décomposer en syntaxèmes et éventuellement raffiner les connexions.

La **prosodie** joue certainement un grand rôle : elle permet, nous pensons, de segmenter une production et d'indiquer pour chacun des segments obtenus qu'il peut être traité indépendamment de la suite et recevoir une structure connexe. La prosodie sert donc (en plus de ses autres fonctions liées à l'expressivité ou au marquage de l'illocution) à guider le traitement de la chaîne parlée par l'interlocuteur.

Prenons un exemple : L'autre jour, j'ai rencontré quelqu'un que j'avais pas revu depuis l'école primaire. Même si l'autre jour est détaché et que le reste de l'énoncé forme une unité plus cohésive (un novau, nous verrons), on n'attend pas d'avoir la totalité du noyau pour connecter *l'autre jour* à la construction verbale. Le complément d'objet direct de RENCONTRER est un groupe substantival assez long : quelqu'un que j'avais pas revu depuis l'école primaire. Même s'il faut avoir traité l'ensemble de ce groupe pour savoir 'ce que j'ai rencontré', on n'a pas besoin d'attendre d'avoir traité la relative pour connecter quelqu'un à j'ai rencontré. Il est d'ailleurs assez probable que l'énoncé soit segmenté prosodiquement de la façon suivante : l'autre jour | j'ai rencontré quelqu'un | que j'avais pas revu | depuis l'école primaire, puisque les groupes accentuels tendent à faire autour de 6 syllabes, et qu'il faut que, à chaque frontière de groupe accentuel (marquées ici par |), l'ensemble de ce qui précède reçoive une analyse syntaxique complète. Chacun des quatre groupes prosodiques peut recevoir une analyse complète indépendamment des autres groupes, puis être connecté aux autres, ce qui donne quelque chose comme ça:

Nous pouvons dire pour conclure que la représentation de l'organisation syntaxique d'un énoncé par une structure de connexions présente des avantages pour qui voudrait étudier plus avant le traitement cognitif de la parole et de la lecture, permettant de rendre compte d'une connexion à différents niveaux de granularité, d'attribuer des analyses à n'importe quelle portion d'un énoncé et de construire la structure d'un énoncé de manière incrémentale.





Exercices

Exercice 1 a) Expliquez quelles contraintes agissent sur la réalisation du sens 'rapide' dans les deux phrases suivantes :

- (23) Zoé a jeté un coup d'œil **rapide** dans la pièce.
- (24) Zoé a regardé rapidement dans la pièce.
- b) Quelles contraintes agissent sur la réalisation de 'épée' dans les deux phrases suivantes?
- (25) Manier une épée peut être dangereux.
- (26) Le maniement d'une **épée** peut être dangereux.

Exercice 2 Comparer les paradigmes de commutation de la position sujet du verbe plaire et de la position objet du verbe aimer. Quelles sont les différences?

- (27) Le chocolat plaît à Jean.
- (28) Jean aime le chocolat.

Exercice 3 Donner la liste des unités syntaxiques autonomisables (c'està-dire plus grandes ou égales à un mot) de l'énoncé :

Cet étudiant veut acheter un livre de syntaxe

Exercice 4 En utilisant le test d'autonomisabilité illocutoire, déterminer les unités syntaxiques autonomisables des énoncés suivants et en déduire la structure de connexion correspondante.

- a) Ca a commencé bien des années plus tard.
- b) le plus haut mur du monde.

Exercice 5 On propose la structure de connexion suivante pour *Le chat dort sur le lit*.

Cette structure est donnée sous forme de polygraphe (voir encadré ??).

- a) Quelles sont les unités qu'on peut déduire de cette structure?
- b) À quels critères répond cette structure de connexion?
- c) Peut-on raffiner les connexions de cette structure ? En utilisant quels critères ?



Lectures additionnelles

L'organisation de nos chapitres doit beaucoup aux travaux d'Igor Mel'čuk. Celui-ci considère 3 types de critères pour définir la structure syntaxique : les critères A décident si deux mots se combinent entre eux, les critères B lequel des deux mots est la tête du syntagme qu'ils forment, les critères C quelle est la relation qui les unit. Le présent chapitre est ainsi consacré aux combinaisons et donc aux critères A. Nous avons étendu l'analyse de Mel'čuk en ne limitant pas la connexion aux combinaisons de mots. Les critères B seront étudiés au Chapitre 9 sur *Tête et dépendance*. Les critères C seront abordés plus tard, dans le Chapitre ?? sur *Les relations syntaxiques*.

L'autonomisabilité prosodique et plus généralement la relation entre une structure de dépendances syntaxique a été étudiée par Piet Mertens dans sa thèse en 1987. Pour un travail plus récent, on pourra consulter son article de 2008. Pour un travail précurseur sur les liens entre prosodie et syntaxe de dépendance, citons le travail d'Henri Weil1844 sur lequel nous reviendrons.

Lucien Tesnière utilise des bulles dans ses représentations syntaxiques. Une formalisation des arbres à bulles est proposée par **Kahane1997**. La



définition de la structure de connexion par raffinement d'une fragmentation est proposée par Gerdes et Kahane2011. La définition de la connexion comme une classe d'équivalence de combinaisons est développée dans Kahane 2018. Le lien historique entre connexion et analyse en constituants immédiats, chez Henry A. Gleason comme chez ses prédécesseurs, est étudié par Mazziotta et Kahane2017. Timothy Osborne et ses collègues ont consacré un article aux catenae en 2012.

Sur le traitement cognitif de la syntaxe, on consultera les travaux précurseurs de Janet D. Fodor et de son étudiante Lyn Frazier. Dans leur article de 1978, elles défendent l'idée que les énoncés sont traités par petits paquets d'environ six mots qui sont ensuite connectés entre eux. Cette étude est suivie en 1982 et 1994 d'une analyse du traitement des phrases avec garden-path par l'analyse des mouvements oculaires, où il est défendu que l'énoncé n'est pas réanalysé, mais que le destinataire répare sa première analyse après avoir identifié la source de l'erreur.

Janet D. Fodor, Atsu Inoue1994 The diagnosis and cure of garden paths, *Journal of Psychologuistic Research*, 23(5), 407-434.

Lyn Frazier, Janet D. Fodor1978 The sausage machine: A new two-stage parsing model, Cognition, 6(4), 291-325.

Lyn Frazier, Keith Rayner1982 Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences, Cognitive Psychology, 14, 178-210.

Kim Gerdes, Sylvain Kahane2011 Defining dependency (and constituency), Proceedings of the 1st international conference on Dependency Linguistics (Depling), 17-27.

Henry A. Gleason1955 An introduction to descriptive linguistics, Holt, Rinehart and Wilston.

Sylvain **Kahane1997** Bubble trees and syntactic representations, *Procee*dings of the 5th conference on Mathematics of Language (MoL), 70-76.

Sylvain Kahane 2018 Une approche mathématique de la notion de structure syntaxique : raisonner en termes de connexions plutôt que d'unités, TAL, 59(1).

Nicolas Mazziotta, Sylvain Kahane2017 To what extent is Immediate Constituency Analysis dependency-based? A survey of foundational texts, Proceedings of the Fourth International Conference on Dependency Linguistics (Depling), 116-126.

some lists of references are firstname, lastname; others are lastname, firstname. Unify.

Igor Mel'čuk1988 Dependency Syntax : Theory and Practice, State University of New York.

Piet **Mertens1987** *L'intonation du français : de la description linguistique à la reconnaissance automatique*, thèse de doctorat, Katolieke universiteit te Leuven.

Piet Mertens2008 Syntaxe, prosodie et structure informationnelle : une approche prédictive pour l'analyse de l'intonation dans le discours, *Travaux de linguistique*, 56(1), 97-124.

Timothy Osborne, Michael Putnam, Thomas **Gross2012** Catenae : Introducing a novel unit of syntactic analysis, *Syntax*, 15(4), 354-396.

Henri **Weil1844** *De l'ordre des mots dans les langues anciennes comparées aux langues modernes*, thèse de doctorat, Paris.



Citations originales

Citations de la section 8.25. Gleason19551961 :

We may, as a first hypothesis, consider that each of [the words of the considered utterance] has some statable relationships to each other word. If we can describe these interrelationships completely, we will have described the syntax of the utterance in its entirety. [...] We might start by marking those pairs of words which are felt to have the closest relationship. [...] We will also lay down the rule that each



word can be marked as a member of only one such pair.



Corrections des exercices

Corrigé 1 a) Dans le premier cas, rapide se combine avec la locution nominale coup d'œil, ce qui explique la réalisation comme adjectif. Dans le deuxième cas, la combinaison a lieu avec le verbe REGARDER, ce qui entraine la réalisation par l'adverbe rapidement.

b) Dans le premier cas, une épée est le complément d'objet direct de la forme infinitive du verbe MANIER. Dans le deuxième cas, une épée est le complément du nom maniement; dans ce cas, une épée doit d'abord se combiner avec la préposition DE avant de se combiner avec *maniement*.

Corrigé 2 Les deux positions acceptent des groupes substantivaux, des infinitives (*lire le soir*) et des propositions complétive (*qu'il pleuve le matin*). On observe deux différences dans les paradigmes. D'une part, les pronoms sujets et objets sont différents : il plaît à Marie vs Marie l'aime. D'autre part, la position sujet accepte des infinitives en de : De lire le soir plaît à Marie vs *Marie aime de lire le soir.

Corrigé 3 Les unités syntaxiques autonomisables de cet énoncé sont :

- les 8 mots : cet, étudiant, veut, acheter, un, livre, de, syntaxe;
- 4 segments de deux mots : cet étudiant, veut acheter, un livre, de syntaxe:

- 3 segments de trois mots : cet étudiant veut, acheter un livre, livre de syntaxe;
- 3 segments de quatre mots : cet étudiant veut acheter, veut acheter un livre, un livre de syntaxe;
- 1 segment de cinq mots : acheter un livre de syntaxe;
- 2 segments de six mots : cet étudiant veut acheter un livre, veut acheter un livre de syntaxe;
- l'énoncé complet : cet étudiant veut acheter un livre de syntaxe.

Corrigé 4 a) bien des années plus tard forme un syntagme, puisqu'il peut être la réponse à la question « Ca a commencé quand? ». À la même question, on peut répondre des années plus tard, bien plus tard ou plus tard, voire juste tard, mais pas bien des années, bien, des années ou plus. On en déduit que bien et des années se combinent avec plus tard. Il est plus difficile de déterminer si bien et des années se combinent avec plus ou tard, puisque ni des années plus, ni des années tard n'est acceptable. On peut néanmoins remarquer qu'on trouve un groupe substantival avec plus dans d'autres contextes: Il habite un étage plus haut; Fais-le deux centimètres plus large. Par contre, on ne peut jamais combiner un groupe substantival avec tard sans la présence de plus, ce qui nous permet de conclure que des années se combine avec plus. Si bien peut se combiner aussi bien avec plus que tard, il semble, au vu du sens, que c'est avec plus qu'il se combine ici.

b) La fragmentation du syntagme *le plus haut mur du monde* pose plusieurs problèmes. Nous avons déjà traité le cas des adjectifs épithètes et expliqué pourquoi *le haut* n'est pas un fragment acceptable. Mais c'est plus compliqué avec *le plus haut*, puisqu'on a *le mur le plus haut*. On peut argumenter que *le* dans *le plus haut mur* est bien l'article, car il commute avec un possessif : *son plus haut mur*, *mon plus vieil ami*. Ce qui n'est pas le cas qu'en *le plus haut* forme une unité syntaxique et un groupe adjectival postposé au nom : **le mur notre plus haut*. On aurait donc la structure de connexion suivante pour *le plus haut mur* :

Mais on peut aussi accepter que dans cette construction le joue un double rôle : à la fois article défini ($le\ mur$) et marqueur du superlatif ($le\ plus$), ce qui donne une structure cyclique.

Un autre problème d'analyse vient du modifieur du monde. Il ne s'agit



pas d'un complément du nom *mur*, puisque #*mur du monde*, s'il est acceptable, n'a pas le sens requis et surtout on a le mur le plus haut du monde. De même, du monde n'est pas un modifieur de l'adjectif, car l'absence du superlatif donne un sens inapproprié : #le haut mur du monde. Donc, bien que le fragment le plus du monde soit inacceptable, nous pouvons considérer que la présence du modifieur du monde est validé par la présence du superlatif et que les deux sont donc connectés.

Corrigé 5 a) Les unités induites sont les mots, ainsi que le lit, sur le lit, le chat, le chat dort, dort sur le lit et la phrase entière.

- b) Ces unités obéissent aux critères d'autonomisabilité. En particulier, sur ne peut apparaître sans son complément le lit.
- c) On peut raffiner la structure en utilisant le critère distributionnel. Nous verrons au chapitre suivant qu'il y a de bonnes raisons de considérer la combinaison $dort \oplus sur$. Il est moins facile de raffiner les connexions mettant en jeu une unité déterminant-nom comme le chat ou le lit.

9 Tête et dépendance : Hiérarchiser la structure syntaxique

9.1 Hiérarchie

Les syntaxèmes, mots et autres unités syntaxiques ne sont pas seulement connectés les uns aux autres : certains dominent les autres et leur imposent leurs propriétés syntaxiques (catégorie, forme, fonction, place). Nous avons largement argumenté l'existence d'une structure syntaxique hiérarchique lorsque nous avons étudié les exemples *Pierre a été malade pendant deux semaines* et *La maladie de Pierre a duré deux semaines* dans le Chapitre 2. Nous renvoyons le lecteur à cette discussion. De nouveaux éléments seront donnés dans la suite de ce chapitre. En particulier :

- les contraintes distributionnelles;
- la rection;
- la hiérarchie sémantique;
- les contraintes sur l'ordre des mots.

Avant cela, nous allons introduire un peu de terminologie, en nous basant sur la notion de connexion, introduite au chapitre précédent.

9.2 Tête et dépendance

La hiérarchie de la structure syntaxique se traduit par une asymétrie des combinaisons entre unités, lesquelles combinent alors une unité hiérarchiquement supérieure à une unité qui lui est assujettie.

Une connexion hiérarchisée est appelée une dépendance.

Si A et B sont connectés et que A est hiérarchiquement supérieur à B, on dit que A gouverne B et que B dépend de A. Ou encore que A est le gouverneur de B et que B est un dépendant de A. En général, un élément peut avoir un nombre quelconque de dépendants, mais il n'a qu'un seul gouverneur.

En syntaxe de dépendance, depuis les travaux fondateurs de Lucien Tesnière, les dépendances syntaxiques lient des mots entre eux. Dans notre approche, les dépendances, comme les connexions, peuvent être considérées entre n'importe quelles unités syntaxiques et une même dépendance peut être décrite de façon plus ou moins fine en l'attribuant à des unités syntaxiques plus ou moins larges (voir la section 8.14 sur *La connexion et ses instances*). Nous manipulerons des structures de granularité variable : parfois les mots seront les nœuds de la structure, mais parfois nous considérerons des unités plus fines (les syntaxèmes) ou des unités plus larges (voir la section 8.18 sur *Structures de connexion, granularité et critères* et la section 10.2 sur l'*Arbre de Beauzée-Gladkij*).

Une dépendance est représentée par une **flèche** qui va **du gouverneur vers** le dépendant ou encore simplement en positionnant le gouverneur au dessus du dépendant. Nous illustrons cela avec l'énoncé *Marie parle* où le verbe *parle* gouverne *Marie*.

Figure 9.1 - Deux représentations d'une dépendance

Le sens de la flèche, du gouverneur vers le dépendant, est une convention. Comme toute convention, elle est en partie arbitraire. Il aurait été tout à fait possible d'orienter les flèches dans l'autre sens, comme certains auteurs l'ont fait. Nous adoptons ici la convention qui est la plus largement répandue.

La représentation avec le gouverneur au dessus est également arbitraire, même si l'on est largement habitué aujourd'hui à une telle représentation de la hiérarchie (voir par exemple l'organigramme des responsabilités dans une entreprise). Dans ses premiers schémas en 1934, Tesnière adoptait une autre convention : il plaçait l'élément dominant au centre comme le soleil au centre du système solaire (voir l'encadré ?? sur l'Historique des représentations syntaxique par des diagrammes en dépendance).

Nous allons introduire un autre terme à ne pas confondre avec *gouverneur*.

On appelle TÊTE d'une unité syntaxique U toute sous-unité de U qui n'est gouvernée par aucune autre sous-unité de U.

Les notions de tête et de gouverneur renvoient au même concept, mais adopte des points de vue différents.

Si A **gouverne** B, alors A et B forment une unité syntaxique ensemble et A en est la **tête**. Inversement, si A et B forment à eux deux une unité dont A est la **tête**, alors B est **dépendant** de A.

Autrement dit, hiérarchiser une connexion revient donc à décider de quel côté est sa tête.

Figure 9.2

On aura noté que pour une unité syntaxique U donnée, la tête de U est un élément de U, en quelque sorte l'élément le plus important de U du point de vue de la syntaxe, tandis que le gouverneur de U est un élément extérieur à U. (On peut se souvenir, pour ne pas confondre les deux termes, que la tête d'une personne fait toujours partie de cette personne.) Illustrons ces notions avec *une très jolie valise* et U = *très jolie*. La tête de U est *jolie*, tandis que le gouverneur de U est *valise*.

Figure 9.3 - Gouverneur vs. tête

Les grammaires de dépendance traditionnelles font l'hypothèse que toute unité possède un unique mot tête. Nous considérons pour notre part que la nature de la tête dépend de la granularité de l'analyse et qu'on peut considérer, selon les finalités de l'analyse, une unité plus fine que le mot (un syntaxème) comme une unité plus large. Par ailleurs, il peut arriver que plusieurs éléments possèdent des propriétés de tête : on parle alors de CO-TÊTES (voir les sections 3.3.23 à 3.3.27 sur *Nom ou déterminant comme tête*).

Concluons cette section en soulignant une propriété fondamentale que nous exploiterons pour identifier la tête d'une unité.

Le gouverneur d'une unité U est connecté à la tête de U.

9.3 Historique des notions de dépendance et de tête 🕙

Le premier grammairien à utiliser le terme « dépendance » dans un sens grammatical est semble-t-il un grammairien arabe, Ibn Mada, qui vécut en Espagne (alors sous domination arabe) entre 1119 et 1195. Il est connu pour son livre, intitulé La réfutation des grammairiens, dans lequel il s'attaque à des sujets qui sont toujours d'actualité. Il critique l'utilisation de formes invisibles sous-jacentes et notamment l'introduction d'un morphème zéro pour le nominatif. Il se prononce aussi en faveur d'une indépendance de la syntaxe et de la sémantique et contre une justification sémantique ou une interprétation cognitive des règles grammaticales : « Et pourquoi l'agent est-il au nominatif? La réponse correcte est [...] que c'est ainsi que parlent les Arabes. » Comme nous l'avons dit, il utilise le terme XXXX ta'alluq, qui se traduit par être accroché à, dépendre de, être connecté à, joindre, attachement, amour du monde, dépendance, connexion, relation, et même obsession. Il préfère ce terme à \sum amal 'opération, gouverner', le terme utilisé communément à son époque en grammaire pour décrire des relations entre le gouverneur et le dépendant. Selon Ibn Mada, le gouverneur n'opère pas sur ses dépendants, il y a seulement une relation, une dépendance. Il va même jusqu'à appeler hérétique toute utilisation de ^camal car, selon lui, des mots ne peuvent agir sur d'autres mots et déclencher une flexion. Comme l'utilisation de ta'alluq était essentiellement un changement terminologique pour une notion établie de longue date sous le terme de ^camal, le terme de dépendance n'a pas vraiment pris avant le 20^e siècle et les travaux de Lucien Tesnière, qui lui donnaient une assise théorique forte (bien que Tesnière lui-même utilise le terme de connexion).

L'importance d'une analyse de type dépendentiel pour la description de phénomènes grammaticaux, par contre, était établie bien avant Ibn Mada. On attribue généralement à **Sibawayh**, le grand grammairien perse travaillant sur l'arabe, la première modélisation grammaticale en termes de dépendances entre mots et notamment en ce qui concerne l'attribution des cas aux noms dépendant d'un verbe.



Sibawayh a vécu de 760 à 796 et on raconte qu'il est mort, très jeune, de la fureur qui l'a saisi lors d'un débat sur la grammaire de l'arabe (premier accident de travail d'un linguiste [F04C?]). Il est le fondateur d'une tradition grammaticale qui reste aujourd'hui vivace.

Avant l'avènement de la tradition grammaticale arabe, on analysait déjà la structure des phrases en faisant référence à des liens hiérarchisés entre les mots. La plus vieille grammaire étendue qu'on connaît, les descriptions du sanskrit par l'immense linguiste indien **Pānini** au 4^e siècle avant notre ère, introduit la notion de karaka pour désigner les relations entre verbes et noms, qui sont classées en six classes différentes en se basant sur des critères sémantiques autant que syntaxiques. Parmi ces classes de liens, on trouve déjà l'agent/sujet (karta) et le patient/objet (karma). Il faudra de toute façon attendre le 20e siècle pour une distinction claire entre les notions syntaxiques de sujet et objet d'une part et les notions sémantiques d'agent et de patient d'autre part (voir l'encadré sur Sujet, agent, thème au Chapitre ??).

Généralement, on constate qu'une analyse en termes de dépendances s'est imposée pour pratiquement toute analyse grammaticale jusqu'à l'apparition des grammaires de constituants au 20^e siècle, en particulier pour des langues flexionnelles et à ordre libre, comme le sanskrit, l'arabe ou encore l'allemand.

L'idée d'une analyse syntaxique complète d'un énoncé apparaît, à notre connaissance au début du 18^e siècle, en France. Il est probable que l'impulsion donnée au 17^e siècle par la parution de la *Grammaire de Port-Royal* (ArnauldLancelot1660), puis de la Logique de Port Royal (ArnauldNicole1662) a été déterminante. En 1709 paraît la Grammaire françoise sur un plan nouveau de Claude Buffier qui comprend plusieurs analyses très élaborées. Malgré le caractère très moderne de ses analyses, Buffier reste relativement peu connu, probablement en raison d'une terminologie qui ne distingue pas clairement les relations syntaxiques de la morphologie. Si l'on remplace, dans les citations qui suivent, régime par complément ou dépendant, on voit que les analyses de Buffier sont de parfaites analyses en dépendance : « Tous les noms ou même tous les mots qui servent ainsi à particulariser la signification d'un autre mot, sont le régime de ce mot : comme si je dis un ami de plaisir, la signification d'un ami est particularisée par le mot de plaisir; c'est pourquoi de plaisir est le régime d'un ami. » (p. 57) « Dieu agit avec justice. Avec est un mot qui n'a point de sens déterminé et complet par lui-même; mais par le mot justice dont il est ici suivi et qui en est le régime. » (p. 73) Les analyses de Buffier, comme celles qui suivront tout au long du 18^e siècle, sont données de manière discursive (c'est-à-dire sans schéma), ce qui les rend plus difficiles à saisir. Nous nous proposons d'en étudier une en détail dans les exercices à la fin

de ce chapitre (exercice 1, section ??).

Dans son ouvrage de 1747, Les Vrais principes de la langue françoise, ou la Parole réduite en méthode, l'abbé Gabriel Girard utilise le terme dépendance dans un sens compatible avec la définition moderne : « J'abandonne toutes les observations qu'on pourrait faire sur le Style, pour me borner uniquement à ce qui regarde l'union grammaticale des mots. Cette sorte d'union établit entre eux un RÉGIME, qui est très distingué de ce que je viens de nommer style; ce dernier consistant dans des rapports de convenance [= relation d'accord] dont le goût fait choix pour la conduite du discours, et l'autre dans des rapports de dépendance soumis aux règles pour la construction de la phrase. » (p. 87, nous modernisons l'orthographe) Mais Girard préfère néanmoins parler de régime plutôt que de dépendance : « Le Régime n'est autre chose que le concours des mots pour les expressions d'un sens ou d'une pensée. Dans ce concours de mots il y en a qui tiennent le haut bout; ils en régissent d'autres, c'est-à-dire qu'ils les assujettissent à certaines lois : il y en a qui se présentent d'un air soumis ; ils sont *régis* ou tenus de se conformer à l'état et aux lois des autres; et il y en a qui sans être assujettis ni assujettir d'autres, n'ont de lois à observer que celle de la place dans l'arrangement général. » (p. 87) (L'idée, exprimée à la fin de cette citation, qu'il y a des mots qui n'entrent pas dans la rection, sera reprise par l'Approche pronominale et conduira à la distinction entre micro et macrosyntaxe; voir partie 6.)

Nicolas Beauzée, dans l'article « Régime » de l'Encyclopédie (vol. 14, 1765), va introduire la distinction entre la notion de complément et le régime proprement dit (c'est-à-dire « la forme particulière que doit prendre un complément grammatical d'un mot, en conséquence du rapport particulier sous lequel il est alors envisagé »), critiquant au passage l'emploi du terme régime par Girard. Beauzée est ainsi le premier à distinguer clairement dépendance syntaxique et dépendance morphologique (voir encadré suivant). Beauzée ne dessine pas de diagramme, mais il fait une description détaillée de la structure syntaxique qui est similaire aux diagrammes que proposera Gladkij plus de deux siècles plus tard (??), diagrammes que nous appellerons pour cette raison des arbres de Beauzée-Gladkij (voir section 10.2 Arbre de Beauzée-Gladkij) : « Par exemple, dans cette phrase, nous avons à vivre avec des hommes semblables à nous : ce dernier nous est le complément de la préposition à; à nous est celui de l'adjectif semblables; semblables à nous est le complément total du nom appellatif les hommes; les hommes semblables à nous, c'est la totalité du complément de la préposition de. »

Mieux encore, Beauzée distingue le *complément grammatical* ou *initial*, qui est un mot, du *complément logique* ou *total*, qui en est la projection (voir section 9.31 *Dominance et projections maximales*), montrant ainsi le passage d'un arbre de



Beauzée-Gladkij à un arbre de dépendance (que nous détaillerons tout au long du Chapitre 10) : « Par exemple, dans cette phrase, avec les soins requis dans les circonstances de cette nature; le mot nature est le complément grammatical de la préposition de : cette nature en est le complément logique : la préposition de est le complément initial du nom appellatif les circonstances; et de cette nature en est le complément total : les circonstances, voilà le complément grammatical de la préposition dans; et les circonstances de cette nature en est le complément logique. » Il faut néanmoins remarquer que le sujet n'est pas considéré comme un dépendant du verbe, ce que souligne le choix même du terme complément (voir la discussion dans le Chapitre ??). Beauzée énonce même la projectivité (voir la section 11.15 sur la *Projectivité*), notant que les compléments logiques/totaux tendent à être continues.

On peut considérer que Beauzée est le premier linguiste à donner une description claire d'une structure de dépendance. Cette description est faite à base de mots et ne trouve pas encore de traduction graphique, mais elle n'en reste pas moins une véritable analyse en dépendance. Cette opinion converge avec celle de Jean-Claude Chevalier, qui considère dans son monumental ouvrage de 1968 sur l'histoire de la syntaxe de 1530 à 1750 qu'avec Beauzée s'achève la définition de la notion actuelle de complément.

Les idées de Beauzée doivent beaucoup à celle de Dumarsais qui rédige onze ans plus tôt l'article « Construction » de l'Encyclopédie (vol. 4, 1754). Dans cet article, Dumarsais décrit la dépendance en utilisant le terme détermination : « Un mot doit être suivi d'un ou de plusieurs autres mots déterminants, toutes les fois que par lui-même il ne fait qu'une partie de l'analyse d'un sens particulier; l'esprit se trouve alors dans la nécessité d'attendre et de demander le mot déterminant, pour avoir tout le sens particulier que le premier mot ne lui annonce qu'en partie. [...] Quelqu'un me dit que le Roi a donné; ces mots a donné ne font qu'une partie du sens particulier, l'esprit n'est pas satisfait, il n'est qu'ému, on attend, ou l'on demande, 1° ce que le Roi a donné, 2° à qui il a donné. On répond, par exemple, à la première question, que le Roi a donné un régiment : voilà l'esprit satisfait par rapport à la chose donnée; régiment est donc à cet égard le déterminant de a donné, il détermine a donné. On demande ensuite, à qui le Roi a-t-il donné ce régiment? on répond à monsieur N. ainsi la préposition à, suivie du nom qui la détermine, fait un sens partiel qui est le déterminant de a donné. » (Voir exercice 2, section ?? pour une autre analyse de Dumarsais.)

Sans que le terme *complément* ne soit utilisé de facon aussi systématique, Dumarsais ébauche la distinction entre complément grammatical et complément logique qui sera ensuite élaborée par Beauzée et qui préfigure la distinction actuelle entre une analyse en dépendance et une analyse en constituants : « On peut considérer une proposition ou grammaticalement ou logiquement : quand on considère une proposition grammaticalement, on n'a égard qu'aux rapports réciproques qui sont entre les mots; au lieu que dans la proposition logique, on n'a égard qu'au sens total qui résulte de l'assemblage des mots. »

Enfin, on doit également à Dumarsais, l'idée reprise par Tesnière (voir la section 8.8 sur la *Connexion*) que le sens naît des connexions entre les mots : « Il faut d'abord établir comme un principe certain, que les mots n'ont entre eux de rapport grammatical, que pour concourir à former un sens dans la même proposition, et selon la construction pleine; car enfin les terminaisons des mots et les autres signes que la Grammaire a trouvés établis en chaque langue, ne sont que des signes du rapport que l'esprit conçoit entre les mots, selon le sens particulier qu'on veut lui faire exprimer. Or dès que l'ensemble des mots énonce un sens, il fait une proposition ou une énonciation. Ainsi celui qui veut faire entendre la raison grammaticale de quelque phrase, doit commencer par ranger les mots selon l'ordre successif de leurs rapports, par lesquels seuls on aperçoit, après que la phrase est finie, comment chaque mot concourt à former le sens total. » (article « Concordance », *Encyclopédie*, vol. 3, 1753).

L'idée d'une hiérarchie est également développée par l'abbé Roch-Amboise Sicard, qui dans son ouvrage de 1801, Élémens de grammaire générale appliqués à la langue française, présente sa théorie du chiffre : « On désigne certaines prépositions par le chiffre 4, et leur complément, par le chiffre 5. » (p. 106), mais le sujet est considéré comme dominant le verbe : « On désigne le sujet par le chiffre 1, qui est censé être l'expression d'une moitié de l'unité. On désigne la qualité [= l'objet] par le même chiffre, qui représente l'autre moitié de la même unité. On désigne le mot-lien, qui est le verbe, par le chiffre 2. » (p. 105).

La théorie du chiffre de Sicard semble avoir été reprise par le grammairien américain James Brown, puis par le linguiste danois Otto Jespersen, qui en a fait une théorie des rangs, dont le verbe est absent! Jespersen, qui a eu en une grande influence sur ses contemporains, et notamment Tesnière, écrit dans *The Philosophy of Language* en 1924 (Ch. VII): « Dans toute dénomination composite d'une chose ou d'une personne, nous voyons toujours qu'il y a un mot d'importance suprême auquel les autres sont joints ou subordonnés. Ce mot principal est défini (qualifié, modifié) par un autre mot, qui à son tour est défini (qualifié, modifié) par un troisième mot, etc. Nous sommes ainsi amenés à établir différents « rangs » de mots selon leurs relations mutuelles comme défini ou définissant. Dans la combinaison (un) temps extrêmement chaud, le mot temps, qui est évidemment l'idée principale, peut être appelé primaire; chaud, qui définit temps,



secondaire; et extrêmement, qui définit chaud, tertiaire. » Jespersen va appliquer sa théorie des rangs à une analyse formelle d'un grand nombre d'exemples dans Analytic Syntax publié en 1937. Par exemple, It will learn it soon enough 'Il va apprendre ça suffisamment tôt' est analysé S V O 3 4, indiquant ainsi que enough₄ modifie soon3.

L'existence d'une hiérarchie s'est également exprimée à travers la notion de tête (angl. head), dont on peut faire remonter l'introduction au grammairien anglais Henry Sweet (Sweet1891, A new English Grammar, sections 40 and 41): « 40. La relation la plus générale entre des mots dans une phrase est, du point de vue logique, celle de mot-ajout et mot-tête, ou, comme on peut aussi dire, de modifieur et modifié. Ainsi dans les phrases les hommes grands ne sont pas toujours forts, tous les hommes ne sont pas forts, grand, fort et tous sont des mots-ajouts modifiant le sens du mot-tête hommes. De même foncé, rapide, rapidement sont des mots-ajouts dans rouge foncé, il a un pas rapide, il marche rapidement. Stone 'pierre' est un mot-ajout dans stone wall, wall of stone, parce qu'il modife (défini) le sens de wall 'mur'. De même, book (books) est un mot-ajout dans book-seller 'marchand de livres', bookselling, sale of books 'vente de livres', he sells books 'il vend des livres', he sold his books 'il a vendu ses livres', les mots-têtes correspondants étant seller, selling, sells, sold. » Comme on peut le voir dans cette citation suivante, Sweet bascule du concept de tête à celui de gouverneur : « 41. [...] La distinction entre mot-ajout et mot-tête est seulement relative : un même mot peut être un mot-tête dans une phrase ou un contexte et un mot-ajout dans un autre, et un même mot peut être mot-tête et mot-ajout en même temps. Ainsi dans il est très fort, fort est un mot-ajout de il et en même temps mot-tête du mot-ajout très, lequel peut lui-même être un mot-tête, comme dans il est pas très fort. »

L'équivalence entre les notions de tête et de dépendance est formellement montrée par Yves Lecerf dans un rapport de 1960. Cette équivalence est présentée en détail dans le Chapitre 10.

La première formalisation de la notion de tête revient probablement à Leonard Bloomfield. Dans sa grande œuvre, Langage, publiée en 1933 et considérée à juste titre comme l'ouvrage fondateur de la syntaxe de constituants, Bloomfield consacre davantage d'énergie à définir la notion de tête que celle de constituant. Il y distingue notamment les constructions endocentriques (celles qui ont une tête) des constructions exocentriques (celles qui n'en ont pas) : « Toute construction syntaxique met en jeu deux (ou parfois plus) formes libres combinées en un syntagme (phrase), que l'on peut appeler le syntagme résultant. Le syntagme résultant peut n'appartenir à la classe distributionnelle (form-class) d'aucun constituant. Par exemple, John ran n'est ni une expression nominative

(comme John), ni une expression verbale finie (comme ran). Par conséquent, nous dirons que la construction anglaise acteur-action est exocentrique: le syntagme résultant n'appartient à la classe distributionnelle d'aucun constituent immédiat. Inversement, le syntagme résultant peut appartenir à la même classe distributionnelle que l'un (ou plus) des constituants. Par exemple, poor John est une expression nominale propre, comme l'est le constituant John; les formes John et poor John ont, dans l'ensemble, les mêmes fonctions. En conséquence, nous dirons que la construction anglaise caractère-substance (comme dans poor John, fresh milk, etc.) est une construction endocentrique. » On notera que la construction sujet-verbe (appelée par Bloomfield actor-action) est considérée comme exocentrique, une analyse qui s'est perpétuée depuis la Grammaire de Port Royal (ArnauldLancelot1660) et qui va se maintenir dans les grammaires de constituants jusqu'à l'avènement de la syntaxe X-barre à la fin des années 1970 (voir encadré ?? sur la Syntaxe X-barre).

La « CENTRALITÉ DU VERBE » et le fait que la relation sujet-verbe est exocentrique est, elle, défendue par le linguiste hongrois **Brassai** dès 1863 : « Assis au début, au milieu ou à la fin de la phrase, quel que soit l'endroit où il lui plaît d'être, se trouve le verbe, relié par des liens sémantiques à ses vassaux, les dépendants. [...] L'autorité du verbe n'est pas dictatoriale et ses vassaux ne sont pas des esclaves, mais ont des relations légiférées avec leur seigneur et les uns avec les autres ; ils possèdent chacun un certain degré d'autonomie et un certain rang, avec un féodalisme dont le slogan, comme aux temps historiques, est *nulle terre sans seigneur* [en français dans le texte].» (d'après une traduction en anglais du hongrois dans l'article *Constituency or dependency? Notes on Sámuel Brassai's syntactic model of Hungarian* d'András Imrényi en 2013).

Lucien Tesnière, dès son article de 1934, défend avec véhémence la centralité du verbe : « Une phrase se présente comme un système solaire. Au centre, un verbe qui commande tout l'organisme de même que le soleil est au centre du système solaire. » Il insiste sur le fait que le sujet dépend du verbe comme les autres actants, considérant ainsi la combinaison sujet-verbe comme endocentrique : « L'opposition du sujet et du prédicat empêche ainsi de saisir l'équilibre structural de la phrase, puisqu'elle conduit à isoler comme sujet un des actants, à l'exclusion des autres, lesquels se trouvent rejetés dans le prédicat pêle-mêle avec le verbe et tous les circonstants. [...] L'opposition du sujet et du prédicat masque en particulier le caractère interchangeable des actants, qui est à la base du mécanisme des voix actives et passive. » (Tesnières1959, ch. 49) (Le terme prédicat est à prendre ici dans un sens particulier dont nous discuterons à nouveau dans l'encadré Fonctions syntaxiques et prédicat au Chapitre ??.) Dans ses premiers



diagrammes syntaxiques, Tesnière place le verbe au centre (voir encadré ?? Historique des représentations syntaxiques par des diagrammes en dépendance).

Premièrement, Tesnière n'a pas introduit le terme de dépendance. Comme nous l'avons dit au chapitre précédent, Tesnière utilise le terme de *connexion*, une connexion liant toujours pour lui un élément régissant à un élément subordonné. Quant à la structure qu'il associe à une phrase, il l'appelle un STEMMA. Deuxièmement, le stemma n'est pas un arbre de dépendance, mais une structure de dépendance dont certains liens ne sont pas hiérarchisés. À côté de la « connexion », qui est donc une dépendance chez lui, Tesnière considère deux autres types de relations syntaxiques, considérées comme exocentriques : la TRANSLATION et la JONCTION. En conséquence, le stemma n'est pas non plus un graphe, puisqu'en cas de translation, le gouverneur de la translation a la translation complète pour dépendant et non l'un des éléments en relation de translation. Comme nous sommes sur des positions assez proches de Tesnière, nous aurons l'occasion de revenir sur la translation et la jonction et d'expliquer pourquoi il peut être envisagé de ne pas les considérer comme des relations hiérarchiques. La translation est discutée à la section ?? sur La translation syntaxique. Quant à la jonction, elle fait l'objet d'un chapitre entier, le Chapitre ??.

L'idée que toutes les constructions sont endocentriques va être défendue plus tard, dans les années 1960-70, aussi bien du côté de la syntaxe de dépendance avec l'école de Moscou (voir l'ouvrage d'Igor Mel'čuk, Dependency Syntax, de 1988, qui reprend des travaux publiés en russe dans les années 1960-70), que du côté de la syntaxe de constituants avec la syntaxe X-barre (voir l'ouvrage éponyme de Ray Jackendoff de 1977). On est ensuite revenu sur cette position en remarquant que, si l'on peut toujours se donner des critères pour attribuer une tête à chaque connexion (comme le fait Mel'čuk par exemple), il faut néanmoins reconnaître que ces critères ne s'appliquent pas de la même façon avec toutes les constructions et que dans un certain nombre de constructions un deuxième élément peut avoir des propriétés de tête. C'est la position que nous défendrons ici.

Il faut aussi signaler que c'est seulement à partir de la fin des années 1970 que les linguistes vont répertorier de manière systématique les critères qui permettent de caractériser la tête (ou ce qui revient à peu près au même, la dépendance). On trouve des critères syntaxiques explicites pour définir la dépendance dans un article remarquable de Paul Garde de 1977 (voir discussion dans l'exercice 4, section ??) et une tentative de définition systématique de la dépendance dans l'ouvrage d'Igor Mel'čuk de 1988. L'article d'Arnold Zwicky de 1985, intitulé Heads, reste fameux pour avoir répertorié les différents critères de définition

de la tête et montré comment ils coïncidaient ou non selon les constructions. **Richard Hudson** répond à Zwicky dans un article de 1987 en corrigeant certaines analyses et en montrant que les critères sont beaucoup plus convergents que ne le dit Zwicky.

Nous poursuivons cette discussion dans l'encadré ?? consacré à l'Historique des représentations syntaxiques par des diagrammes en dépendance et l'encadré ?? consacré à l'Historique des représentations syntaxiques par des diagrammes en constituants.

9.4 Dépendances morphologique, syntaxique et sémantique

Nous nous intéressons dans ce chapitre à la syntaxe et lorsque nous parlons de dépendance, il s'agit évidemment de DÉPENDANCE SYNTAXIQUE. Igor Mel'čuk a contrasté cette notion avec celles de dépendances morphologique et sémantique.

Les dépendances sémantiques sont les relations prédicat-argument qui existent entre les signifiés des sémantèmes (voir chapitres 2 et ??). Par exemple, un verbe à l'infinitif n'a jamais de sujet, mais il possède toujours un argument sémantique qui peut être restitué par une paraphrase où le verbe est à une forme finie. Par exemple, si on compare Marie promet à Pierre de venir et Marie permet à Pierre de venir, dans le premier cas 'Marie' est l'argument de 'venir' (Marie promet qu'elle viendra) et dans le deuxième cas 'Pierre' est l'argument de 'venir' (Marie permet que Pierre vienne). Dans les deux cas, la dépendance sémantique ne correspond pas à une dépendance syntaxique, car ni Marie de venir, ni à Pierre de venir ne sont des unités syntaxiques. Les dépendances sémantiques seront à nouveau considérées dans le Chapitre ?? sur la Syntaxe profonde, ainsi que dans le Chapitre ?? où nous étudierons les distorsions entre structures syntaxique et sémantique dans les phénomènes d'extraction.

(1) Marie promet à Pierre de venir Marie permet à Pierre de venir Représentations sémantiques

Les dépendances sémantiques doivent être distinguées des RELATIONS ANA-PHORIQUES, qui lient deux éléments, généralement un pronom avec son antécédent comme la théorie du liage et elle dans La théorie du liage soulève autant de problèmes qu'elle n'en résout.

On parle de DÉPENDANCE MORPHOLOGIQUE dès que le « choix » d'un syntaxème flexionnel n'est pas libre, mais dépend du choix d'un autre syntaxème (voir la section 6.11 Collocation et choix liés pour le choix non indépendant d'un



lexème). Cela concerne les phénomènes d'ACCORD (le fauteuil est blanc-Ø vs la chaise est blanche; ce truc est génial-∅ vs ces trucs sont géniaux) et les RÉGIMES (Ich sehe einen Mann 'je vois un homme' vs. Ich helfe einem Mann 'j'aide un homme', avec un complément au datif et non à l'accusatif comme précédemment et qui se traduirait littéralement par 'j'aide à un homme'). Dans le cas de blanc vs blanche, le syntaxème flexionnel d'accord en genre sur l'adjectif BLANC est contrôlé par le genre du nom, c'est-à-dire par le syntactique d'un lexème, alors que dans le cas de génial vs géniaux, le syntaxème flexionnel d'accord en nombre porté par l'adjectif GÉNIAL est contrôlé par le syntaxème de nombre du substantif. Le régime est un syntaxème dont la présence est imposée par un gouverneur à son dépendant; dans nos deux exemples en allemand, le verbe impose à son complément un morphème de cas (cas accusatif pour sehen 'voir' et datif pour helfen 'aider').

On parle de dépendance morphologique seulement quand la contrainte morphologique est incontournable. Cela exclut les cas d'accord d'un pronom anaphorique lorsque la référence du pronom peut être choisie librement. Donnons un exemple:

Amélie veut rendre le livre qu'elle vient d'acheter. (2)

Ici l'interprétation préférée est que elle renvoie à Marie et est donc au féminin parce que Marie est féminin. Mais nous ne considérons pas cela comme une dépendance morphologique, car rien n'empêche d'avoir ici un pronom masculin (même si bien sûr le sens change):

(3)Amélie veut rendre le livre qu'il vient d'acheter.

De même que les dépendances sémantiques, les dépendances morphologiques correspondent généralement à des dépendances syntaxiques. Les régimes et les accords servent précisément à marquer des dépendances syntaxiques (voir la section ?? Un exemple - l'accord - de la description à l'explication). Mais certaines dépendances morphologiques se font néanmoins à distance. Tel est le cas des accords suivants:

Amélie aimerait pouvoir rester belle.

Certains considèrent qu'il n'y a pas en fait de dépendance morphologique à distance, mais que l'accord se fait localement avec un pronom zéro sujet du verbe ou de l'adjectif, lequel est localement lié à un autre pronom et cela de manière récursive jusqu'à la source de l'accord. C'est la Théorie du liage :

Amélie_i aimerait ε_i pouvoir ε_i rester ε_i belle.

Les pronoms zéros sont marqués par des ε et la coréférence par la présence d'un même indice i.

On notera que les principes qui définissent les dépendances morphologiques et sémantiques sont également utilisés (de manière secondaire) pour définir les relations syntaxiques (voir la section 9.19 sur le *Critère sémantique de non-effacement* et l'encadré ?? sur les « *Déterminants complexes* »), en raison de la forte coïncidence de ces notions.

9.5 Structure de dépendance et arbre de dépendance

Au chapitre précédent, nous avons présenté la structure de connexion, qui peut être un graphe ou plus généralement un graphe à bulles ou un polygraphe (voir l'encadré ?? Graphe à bulles et polygraphe). Les connexions sont des liens syntaxiques non hiérarchisés : elles lient deux unités syntaxiques sans attribuer de rôle particulier à l'un ou l'autre.

Une STRUCTURE DE DÉPENDANCE est une structure du même type qu'une structure de connexion, mais dont certains liens syntaxiques sont des dépendances, c'est-à-dire des liens hiérarchisés liant un gouverneur à un dépendant. La structure de dépendance est dite partielle quand au moins une connexion n'est pas hiérarchisée.

Pour que la structure de dépendance représente une hiérarchie, il ne suffit pas que chaque lien soit hiérarchisé. Il faut que la structure possède une autre propriété, l'acyclicité, c'est-à-dire qu'elle ne doit contenir de cycles orientés (voir l'encadré ?? Graphe et arbre). Autrement dit, une structure de dépendance ne contient pas de cycles de dépendances, c'est-à-dire de boucles que l'on peut parcourir de gouverneur en dépendant pour revenir au même point. (Nous verrons une exception au Chapitre ?? lorsque nous étudierons les pronoms interrogatifs et relatifs.)

L'acyclicité n'exclut pas les cycles non orientés de connexions comme on l'a vu dans l'encadré ?? Cycle de connexions.

La présence de cycles non orientés dans une structure de dépendance entraîne que certains nœuds ont deux gouverneurs. (On parle alors de **dag**, *directed acyclic*

9.6 Historique des représentations syntaxiques par des diagrammes en dépendance

Marie commet une faute envers Pierre la montée du nationalisme en Catalogne

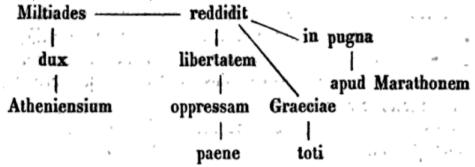
Figure 9.4

graph). Dans cet ouvrage, nous ne considérerons quasiment que des structures sans cycle et pour lesquelles **tout nœud possède au plus un gouverneur**. Plus précisément, lorsque tout nœud a exactement un gouverneur (à l'exception d'un nœud qu'on appelle la racine), la structure est un arbre : l'ARBRE DE DÉPENDANCE.

Notons pour conclure cette section que l'on utilise généralement une terminologie différente lorsqu'on parle d'une unité syntaxique U et de la structure S de dépendance qui lui est associée. Ainsi appelle-t-on RACINE de S le nœud de S qui correspond à la **tête** de U.

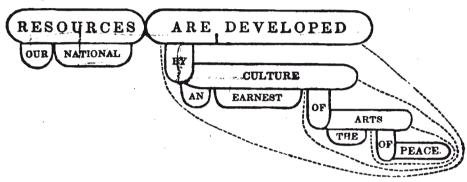
9.6 Historique des représentations syntaxiques par des diagrammes en dépendance

La première représentation dépendentielle que nous connaissons est un diagramme, unique, proposé par le philosophe et grammairien allemand Johann Billroth dans sa grammaire du latin (*Lateinische Syntax*) publiée en 1832, à seulement 24 ans (il décédera à 28 ans).



Stephen W. Clark est le premier grammairien à réellement exploiter des diagrammes dépendentiels dans sa grammaire de l'anglais de 1847 intitulé *The science of the English grammar : A practical grammar in which words, phrases, and sentences are classified to their offices, and their relation to each other, illustrated by a complete system of diagrams* 'La science de la grammaire anglaise : une grammaire pratique dans laquelle les mots, les syntagmes et les phrases sont classées selon leurs fonctions et leurs relations les uns aux autres par un système complet de diagrammes'. Dans les représentations de Clark, les connexions ne sont

pas représentées en tant que telles : un mot qui en « qualifie » un autre (pour reprendre les termes de Clark) est placé dans une bulle sous la bulle de son gouverneur. Les prépositions comme by 'par' ou of 'de' dans la figure ci-dessous reçoivent une forme particulière indiquant qu'elles servent à « connecter » deux mots. Le sujet, le verbe et l'objet direct sont placés au même niveau, comme ici ressources and are developed. Les groupes prépositionnels sont entourés par des bulles en pointillés.



Our national resources are developed by an earnest culture of the arts of peace 'Nos ressources nationales sont développées grâce à une sérieuse culture des arts de la paix'

Les analyses de Clark sont d'une grande cohérence et ses représentations de la coordination ou de l'extraction, sur lesquelles nous reviendrons dans les chapitres 5.3 et 5.4, sont remarquables. Il leur manque juste le développement théorique que proposera plus tard Tesnière.

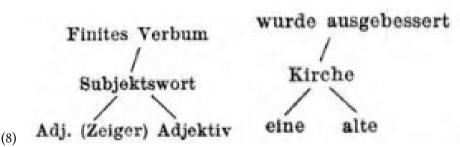
Ce ne sont pas les diagrammes de Clark qui sont passés à la postérité, mais ceux de ses suiveurs, Alonzo Reed and Brainerd Kellogg. Leurs diagrammes, proposés pour la première fois en 1877 et toujours utilisés aujourd'hui par certains enseignants d'anglais, sont équivalents à ceux de Clark, mais utilisent des conventions différentes : les mots ne sont plus placés dans des bulles, mais audessus de segments de traits. Les verbes et les noms reçoivent des segments horizontaux, comme pleased 'enchanté' ou news 'nouvelles' dans le diagramme qui suit, tandis que les autres mots, comme la préposition with 'avec' ou l'adjectif 'good', reçoivent des segments en diagonal. Comme chez Clark, le sujet le verbe et l'objet sont au même niveau. Les connections apparaissent de manière plus explicite à la jointure de deux segments. La connexion entre le sujet et le verbe (girls – are pleased) est symbolisée par un trait vertical et celle entre l'auxiliaire et le participe (are – pleased) par un trait oblique.

(6) The girls are pleased with the good news 'Les filles sont enchantées par la bonne nouvelle' Diagramme de ReedKellogg1877

On peut traduire cette structure dans les conventions utilisées ici, en réifiant les connexions et en mettant en évidence le fait que certaines connexions ne sont pas hiérarchisées :

(7) Interprétation du diagramme de Reed & Kellogg

En 1883, dans un livre sur la grammaire allemande (*Zur Methodik des deutschen Unterrichts* 'Sur la méthodologie de l'enseignement de l'allemand'), Franz Kern propose de véritables arbres de dépendance. Voici ce qu'il écrit : « Le mot déterminant dépend de celui qu'il détermine ou, en d'autres termes, est régi par lui. On désigne (graphiquement, par un schéma) la dépendance d'un mot d'un autre par un trait partant du mot régissant vers le bas et allant vers le mot régi (ou dépendant) [figure de gauche] ou encore sans mot par de simples relations grammaticales [figure de droite]. » Ce texte est accompagné des deux figures suivantes pour la phrase *Eine alte Kirche wurde ausgebessert* 'Une vielle église a été réparée'. Dans la figure de gauche figurent les mots de la phrase, le complexe verbal n'étant pas décomposé. Dans la figure de droite, l'article (appelé *Adj. (Zeiger)* 'adjectif (pointeur)') et l'adjectif qualificatif (*Adjektiv*) dépendent du mot sujet (*Subjektswort*) qui dépend du verbe fini (*Finites Verbum*).

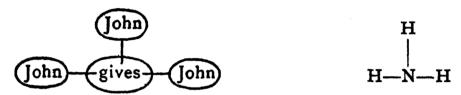


Arbres de dépendance de **Kern1883** *Eine alte Kirche wurde ausgebessert*'Une ancienne église a été reconstruite'

La notion de *valence* a été introduite par le sémioticien anglais Charles S. Peirce dans un article de 1897. Peirce compare la possibilité qu'a le verbe *give* 'donner' de se combiner avec trois éléments (*John gives John to John*) avec la possibilité qu'à l'atome d'azote N de se combiner avec trois atomes d'hydrogène H pour

donner une molécule d'ammoniac NH₃. Cette métaphore de la connexion entre mots par la connexion entre atomes est accompagnée de la figure suivante :

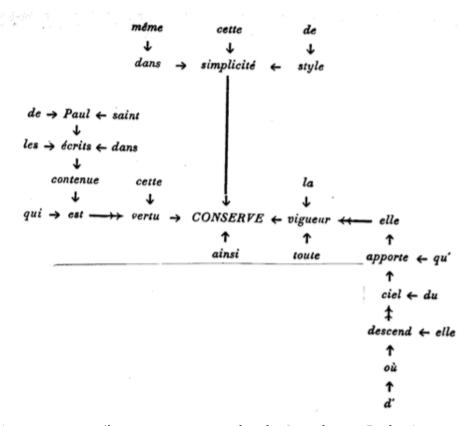
(9) Schéma valenciel de Peirce1897



Peirce remarque que contrairement aux connections chimiques les connections entre mots (il s'agit plutôt de connexions sémantiques que syntaxiques) sont asymétriques, ce que semble confirmer son diagramme, où les connexions partent du mot *gives* mais s'arrêtent à la bulle de *John*. (Voir la section 9.11 sur *Distribution et valence* pour la suite de la discussion.)

C'est à Lucien Tesnière qu'on attribue les bases théoriques de la syntaxe de dépendance, exposées brièvement dans son article de 1934, *Comment construire une syntaxe*, puis en détail dans son ouvrage posthume de 1959, *Éléments de syntaxe structurale*. On notera, dans l'analyse suivante de 1934, que Tesnière considère la préposition comme dépendant du nom qu'elle introduit. (Le diagramme contient par ailleurs une erreur, qui n'est pas dans le manuscrit de Tesnière : le lien entre *apporte* et *vigueur* a été erronément attribué à *elle*.)

(10) Diagramme (tronqué) de Tesnière1934



Ainsi cette vertu céleste, qui est contenue dans les écrits de saint Paul, même dans cette simplicité de style, conserve toute la vigueur qu'elle apporte du ciel d'où elle descend. (Bossuet)

Si cette analyse contient déjà des symboles spéciaux (comme les flèches à double pointe -> pour les relatives), ce n'est que plus tard que Tesnière introduira les symboles en T pour la translation. Les structures de l'ouvrage de 1959 ne sont plus totalement hiérarchique comme le montre notre interprétation polygraphique ci-dessous de la représentation proposée par Tesnière pour *Écrivez dans le livre de votre ami!*.

(11) Stemma de **Tesnière1959** Interprétation polygraphique du stemma

9.7 Arbre et représentation planaire

Il est essentiel de rappeler que la **représentation par un arbre (de dépendance)** sert uniquement à encoder une **structure hiérarchique** liant des nœuds entre eux

par des relations père-fils. Dès qu'un arbre est représenté sur une feuille plane, les nœuds frères (c'est-à-dire ayant le même père) doivent être placés les uns à côté des autres et apparaissent donc comme ordonnés. Cet ordre, imposé par la **représentation planaire**, est souvent utilisé pour ordonner implicitement des arbres (voir le Chapitre 11 où il sera justement question d'ordre linéaire et d'arbres ordonnés). Un arbre ordonné est néanmoins une structure plus riche qu'un simple arbre. Les arbres que nous considérons dans ce chapitre sont des **arbres non ordonnés**. Il faut se les imaginer dans l'espace, libérés de la feuille, comme des **mobiles** suspendus dans l'air et dont les nœuds peuvent tourner librement autour de leur gouverneur.

Pour en faciliter la lecture, nous placerons généralement les nœuds de nos arbres dans l'ordre qu'ils occupent dans la phrase, mais cet ordre n'a aucune pertinence lorsque l'arbre est considéré en tant que tel et nos lecteurs devront essayer d'en faire abstraction. Libérez les arbres de la feuille! Faites-les tourner dans votre tête!

Nous allons maintenant voir comment définir l'arbre de dépendance. Pour cela nous allons introduire différents critères permettant d'identifier la tête d'une unité syntaxique.

9.8 Critère d'effacement simple et critère prosodique

Une première définition possible d'une structure de dépendance se base sur la structure de connexion. Lorsqu'on possède déjà une structure de connexion, on peut la hiérarchiser (au moins partiellement) simplement en choisissant un nœud racine dans la structure. On peut alors orienter les connexions en partant de la racine. Tout se passe comme si on attrapait la structure par ce nœud et qu'on la suspendait.

Illustrons cela avec la phrase:

(12) Pierre veut inviter Marie demain.

Si l'on a la structure de connexion de cette phrase (voir Chapitre 8) et qu'on peut établir que *veut* est la tête (voir section 9.9 suivante), on obtient immédiatement une structure de dépendance par le procédé que nous venons de décrire. La flèche blanche indique le nœud par lequel on attrape la structure pour la hiérarchiser. Ce nœud va gouverner les nœuds auxquels il est connecté, lesquels vont gouverner à leur tour les nœuds auxquels ils sont connectés et ainsi de suite.

Si la structure de connexion n'a que des connexions élémentaires (voir l'encadré ?? Graphe à bulles et polygraphe) et pas de cycles, on obtient un arbre de

Structure de connexion Arbre de dépendance

Figure 9.5 -

dépendance, comme dans l'exemple précédent.

Baser ainsi la dépendance sur la connexion, revient à donner une importance première aux propriétés définitoires des unités syntaxiques, l'AUTONOMISABILITÉ ILLOCUTOIRE et l'AUTONOMISABILITÉ PROSODIQUE (voir la section 8.11 *Unité syntaxique autonomisable*). Traduit pour le repérage de la tête d'une unité, ces deux critères deviennent le critère d'effacement simple et le critère prosodique.

L'autonomisabilité illocutoire donne le critère suivant.

Critère d'effacement simple : si l'unité AB est gouvernée par X et que B peut être effacé, mais pas A, alors A est la tête de AB et B dépend de A.

Dire que B peut être effacé et pas A lorsque AB est gouverné par X revient à dire que XA peut former une unité (illocutoirement) autonome, mais pas XB et donc que X est connecté à A. Comme X est le gouverneur de AB, on en déduit que A est la tête de AB.

Si on reprend notre exemple avec $X = Pierre\ habite$, $A = une\ maison$ et B = écologique et qu'on sait que X gouverne AB, alors on en déduit que B dépend de A, car B est effaçable ($Pierre\ habite\ une\ maison$), mais pas A (* $Pierre\ habite\ écologique$).

Figure 9.6 – Application du critère d'effacement simple

Attention : la présence d'un gouverneur est essentielle dans la formulation du critère d'effacement simple. On trouve souvent la mauvaise formulation suivante : le dépendant d'une connexion est l'élément qui peut le plus facilement être effacé. Cette propriété n'est valable qu'en présence d'un gouverneur. Si AB est un énoncé autonome ou plus généralement si AB n'est pas gouverné on peut très bien effacer la tête de AB. Par exemple, dans Zoé chantait (A = chantait, B = Zoé), on peut effacer A et pas B (Zoé peut former un énoncé autonome, mais pas chantait), alors que c'est A qui est la tête!

Passons maintenant à l'autonomisabilité prosodique, le deuxième critère qui

nous a permis de définir la structure de connexion.

L'autonomisabilité prosodique donne le critère suivant.

CRITÈRE PROSODIQUE : si X est le gouverneur de AB et si X peut former une unité prosodique avec A sans B (sans changer significativement le sens), alors A est la tête de l'unité AB.

Les deux critères que nous venons de proposer sont insuffisants pour plusieurs raisons.

Premièrement, ils ne s'appliquent que dans le cas où l'on étudie une unité qui possède un gouverneur. Ils ne peuvent donc pas être utilisés pour déterminer la tête d'un énoncé. Ce point va être étudié dans la section suivante.

Deuxièmement, il est assez courant que deux éléments qui se combinent soient indissociables, à l'intérieur du mot bien sûr (*chant-ons*), mais aussi en dehors, comme *le chien* dans *le chien dort*. Dans ce cas, le critère d'effacement simple ne s'applique pas, pas plus que le critère prosodique en général.

Troisièmement, le critère d'effacement simple est inopérant si B est un dépendant obligatoire de A. C'est le cas pour une combinaison comme à Marie, où A = à ne peut pas s'employer sans son complément B et ne peut donc jamais former une unité XA avec le gouverneur X de AB. C'est encore le cas pour les formes verbales finies de langues comme le français : en effet, pour Marie dormait, il n'est pas possible de vérifier si A = dormait peut former une unité avec un éventuel gouverneur X, puisque A ne s'emploie jamais sans un sujet B. Ceci a d'ailleurs amené Leonard Bloomfield (qui fut le premier, dans son ouvrage Langage de 1933, à proposer des critères pour définir la tête) à considérer que ces constructions sont exocentriques, c'est-à-dire sans tête (voir la section ?? Historique des notions de dépendance et de tête.)

Les limites du critère d'effacement simple et du critère prosodique nous amèneront à introduire d'autres critères plus puissants. Dans les critères que nous venons de considérer, le gouverneur X de l'unité AB que l'on étudie est fixé. On regarde seulement ce qui se passe dans une phrase donnée, sans faire varier X, A ou B. Cela confère à ces critères une grande simplicité d'utilisation, mais c'est aussi une limite. L'analyse distributionnelle va nous permettre de donner une version plus riche et plus fiable du critère d'effacement : le CRITÈRE DISTRIBU-

TIONNEL AVEC EFFACEMENT, présenté dans la section 9.12 éponyme.

9.9 Tête d'un énoncé

Comme nous l'avons vu à la section précédente, nous avons des critères pour identifier la tête d'une unité dès qu'on connaît son gouverneur, mais ces critères ne peuvent pas s'appliquer pour caractériser la tête d'un énoncé. Or, tant qu'on n'a pas identifié celle-ci et qu'on n'a pas un premier gouverneur, on ne peut pas appliquer le critère d'effacement simple ou le critère prosodique.

L'identification de la tête d'un énoncé repose sur les propriétés de l'énoncé en tant que tel, à savoir le fait qu'il fait l'objet d'une énonciation dirigée vers un interlocuteur et qu'il possède donc une fonction illocutorire. Cette notion repose sur le constat que produire un énoncé est une véritable action de la part du locuteur (on parle d'actes de langage, à la suite des travaux de Gardiner1932, John Austin1962 et John Searle1969), qui attend généralement en retour une action du ou des destinataire(s) (voir également la contribution de Bloomfield1933 dans la section 1.5 Sens et intention communicative).

Ceci nous amène à considérer quatre types d'énoncés : Assertion, Question, INJONCTION et EXCLAMATION.

- assertion : « Il pleut. »; « J'ai mal au ventre. »;
- question : « Comment faire? »; « Est-ce un problème? »;
- injonction : « Laissez ça ici! »;
- exclamation : « Aie! »; « Comme c'est sympa! ».

Les assertions peuvent être acceptés ou refusées par le destinataire; elles se caractérisent par le fait de pouvoir être falsifiées par « *C'est faux.* » ou au contraire validées par « *C'est vrai.* ». Par exemple, si quelqu'un nous dit « *J'ai mal au ventre.* », on peut lui répondre « *C'est faux.* », mais s'il dit « *Aie!* », ce n'est plus possible. De même, on ne peut répondre « *C'est faux.* » a une question ou une injonction. Les questions attentent une réponse. Les injonctions attendent généralement un acte non verbal. Les exclamations attendent simplement d'être partagées par le destinataire éventuel.

La fonction illocutoire a souvent un marquage spécifique. Les éléments qui servent à marquer la fonction illocutoire seront appelés des MARQUEURS ILLOCUTOIRES.

Nous faisons l'hypothèse que les **marqueurs illocutoires** sont **portés** par la TÊTE SYNTAXIQUE DE L'ÉNONCÉ et la caractérise.

En français, le principal marqueur illocutoire est la prosodie (transcrit à l'écrit par des signes de ponctuation comme «? » ou «! »). Mais pour les énoncés à tête verbale en français, il existe d'autres marques. Voyons sur un exemple :

(13) Pierre a dormi.

Cet énoncé est une assertion. Comme toute assertion, on peut la nier en répondant « *C'est faux* ». La négation de cet énoncé ('il est faux que Pierre a dormi') peut s'exprimer en français par un double marquage *ne...pas* qui semble bien identifier l'auxiliaire comme la tête :

(14) Pierre n'a pas dormi.

On peut également faire varier la fonction illocutoire et transformer cet énoncé en une question ou une injonction :

(15) Pierre a-t-il dormi?

(16) Ais dormi (quand je reviens)!

Le français a pour cela des formes particulières : une question peut prendre la FORME dite INTERROGATIVE, qui se caractérise par la présence d'un ENCLITIQUE (-t-il dans notre exemple); une injonction peut s'exprimer par une forme verbale particulière : l'IMPÉRATIF (ici ais, forme impérative du verbe AVOIR). C'est encore une fois l'auxiliaire qui est touché, ce qui confirme son statut de tête.

On peut par la même méthode identifier la tête d'une phrase complexe avec deux verbes finis. Cette tête est traditionnellement appelée le VERBE PRINCIPAL. Considérons :

(17) Marie pense que Pierre dort.

La forme interrogative est *Marie pense-t-elle que Pierre dort*? et pas **Marie pense que Pierre dort-il*? On en déduit que *pense* est la forme verbale principale et que *dort* est une forme verbale subordonnée.



La subordination est simplement le terme traditionnel pour désigner la dépendance d'une forme verbale à un autre élément. Une unité syntaxique dont la tête est une forme verbale est appelée une PROPOSITION. Une proposition dont le verbe est subordonné est appelée une PROPOSI-TION SUBORDONNÉE.

L'identification de la subordination nous procure un nouveau test pour repérer la tête d'une proposition (voir critère rectionnel à la section 9.17 Tête interne et critère rectionnel). En effet, certains verbes imposent un mode particulier à leur subordonné. Par exemple, si l'on subordonne au verbe FALLOIR la proposition Pierre a dormi, on obtient Il faut que Pierre ait dormi (quand je reviens). Comme on le voit, encore une fois, c'est l'auxiliaire qui hérite ici de la marque de subordination, à savoir le MODE SUBJONCTIF (ait est une forme subjonctive de AVOIR), ce qui confirme son statut de tête syntaxique de la proposition.

On peut encore utiliser ces différents critères pour repérer que certains énoncés n'ont pas une tête verbale. Comparons les deux phrases suivantes :

- (18)Heureusement, Pierre a dormi.
- (19)Heureusement que Pierre a dormi.

On peut subordonner le premier énoncé, mais pas le deuxième :

- (20)Je crois que, heureusement, Pierre a dormi.
- (21)*Je crois qu'heureusement que Pierre a dormi.

Comme il apparaît par ailleurs que le verbe croire peut subordonner n'importe quelle proposition à l'indicatif, on peut supposer que la phrase introduite par heureusement que ne peut pas être subordonnée car le verbe n'en est pas la tête. On en déduit que heureusement est la tête de la phrase et subordonne la proposition Pierre a dormi (ce que confirme l'emploi de la conjonction de subordination que).

Le même type de raisonnement permet d'identifier la tête d'une proposition dans n'importe quelle langue a priori (voir encadré ci-dessous).

9.10 Tête d'un énoncé dans une langue inconnue 🕏



Comme nous l'avons dit dans les sections précédentes, l'identification de la tête d'un énoncé est un préalable à la hiérarchisation de la structure.

Nous allons considérer le cas de deux langues non apparentées.

- 1. Commençons par l'anglais, qui est une langue très proche du français, mais dont le fonctionnement est quand même différent. Considérons l'énoncé Peter has slept 'Pierre a (déjà) dormi'. Comme en français, la négation va sur l'auxiliaire sur lequel elle se cliticise :
 - Peter hasn't slept. 'Pierre n'a pas dormi.' (22)

Quant à la question, elle est normalement formée par l'antéposition de l'auxiliaire:

Has Peter slept? 'Pierre a-t-il dormi?' (23)

On peut considérer que cette place particulière de l'auxiliaire le marque comme la tête. D'ailleurs, un verbe ordinaire ne peut recevoir directement la négation, ni être antéposé. Le passage à une forme interrogative ou négative nécessite l'introduction de l'auxiliaire po 'faire':

- (24)Peter slept. 'Pierre a dormi.'
- Did Peter sleep? 'Pierre a-t-il dormi?' (25)
- Peter didn't sleep. 'Pierre n'a pas dormi.' (26)

Comme on le voit, la forme passé slept passe à la forme infinitive. L'auxiliaire lui-même hérite de la finitude (did est la forme passé de DO) et constitue donc la tête de la phrase quand il est présent. En l'absence d'auxiliaire, c'est la forme verbale simple qui est la tête, puisque c'est elle qui est affectée lorsque l'auxiliaire est introduit. On notera que l'auxiliaire, bien qu'il soit la tête de la phrase, peut dans certains cas se cliticiser sur le sujet (c'està-dire perdre la possibilité de recevoir un accent tonique et former un mot prosodique avec le dernier mot du sujet):

(27) Peter's sleeping = Peter is sleeping 'Pierre est en train de dormir', lit. Pierre est dormant.

- (28) Peter's eaten = Peter has eaten 'Pierre a (déjà) mangé'
- 2. En coréen, le verbe principal d'un énoncé est caractérisé par un marqueur de politesse qu'il est le seul à pouvoir porter. Cette particule s'ajoute optionnellement sur le verbe principal, mais jamais sur un verbe subordonné. Le coréen utilise essentiellement des constructions à verbe support (voir Chapitre ??) avec le verbe hada 'faire'. Dans la phrase suivante, le sens 'aimer' est ainsi lexicalisé par le nom sarang 'amour' et le sens 'croire' par le nom saenggak 'pensée', tous deux « verbalisés » par le verbe hada. La deuxième occurrence de hada peut porter le morphème de politesse yo, mais pas la première. (Nous reviendrons sur les marqueurs illocutoires du coréen dans l'encadré ?? Des prédicatifs uniquement locutifs assertifs.)
 - (29) a. Ch'ŏlsu-ga na-rŭl sarang-ha-ndago Ch'ŏlsu-nom moi-ACC amour-faire-PRES.INF saenggak-hae(-yo) pensée-faire(-POLITESSE) 'Je crois que Cholsu m'aime'
 - b. * Ch'ŏlsu-ga na-rŭl sarang-hae-yo-ndago Ch'ŏlsu-nom moi-ACC amour-faire-PRES.INF saenggak-hae(-yo) pensée-faire(-POLITESSE)

9.11 Distribution et valence

En nous basant sur la structure de connexion et en déterminant la tête de la proposition grâce aux marqueurs illocutoires, nous avons vu que nous pouvions obtenir une structure de dépendance partielle. Nous allons voir que l'on peut raffiner cette structure en ajoutant des critères additionnels. Les premiers de ces critères sont les critères distributionnels que nous présenterons dans les sections suivantes. Nous devons avant cela introduire les notions de distribution (voir la section 4.8 sur *L'identification des unités de la langue*, où il a été question d'analyse distributionnelle) et valence (voir la section ?? *Historique des représentations syntaxiques par des diagrammes en dépendance* pour un début de discussion).

La distribution d'une unité est l'ensemble des contextes ou environnements où peut se trouver cette unité.

Deux remarques essentielles sont à faire concernant la distribution.

Premièrement, nous faisons de l'analyse distributionnelle sur un objet structuré. Ici, ce qui nous intéresse, c'est la distribution syntaxique et non la distribution linéaire. Nous nous intéressons donc au contexte des unités au sein de la structure syntaxique, c'est-à-dire au paradigme des unités qui se combinent avec l'unité que nous étudions.

La notion de distribution est encore exprimable en termes de valence.

La valence d'une unité syntaxique est la capacité qu'à cette unité à se combiner avec d'autres unités syntaxiques. La valence est avant tout une valence potentielle, à contraster avec la valence réalisée, lorsque l'unité est utilisée dans un contexte donné.

Cette métaphore de la valence est inspirée de la chimie et de la capacité qu'à un atome à se combiner avec d'autres atomes pour former des molécules (voir la contribution de Peirce1897 dans l'encadré ?? Historique des représentations syntaxiques par des diagrammes en dépendance, ainsi que les formules utilisées par Jespersen1937, évoquées dans l'encadré ?? Historique des notions de dépendance et de tête). Mais alors que les connexions entre atomes sont symétriques (chaque atome fournit un électron de même nature), les connexions entre unités sont doublement asymétriques : asymétrie syntaxique, puisqu'elles lient un gouverneur à un dépendant, et asymétrie sémantique, puisqu'elle lient un prédicat à un argument.

L'asymétrie syntaxique nous amène à distinguer deux parties dans la valence.

La VALENCE SUPÉRIEURE d'une unité syntaxique U est la **position syntaxique** occupée par le **gouverneur** de U, cette position étant elle-même caractérisée par le **paradigme** des éléments qui peuvent l'occuper.

A l'inverse, la VALENCE INFÉRIEURE d'une unité syntaxique est l'ensemble des positions syntaxiques qui **dépendent** potentiellement de cette unité.

Les notions de valences supérieure et inférieure ont été introduites par Igor Mel'čuk, qui les nomment *valence passive* et *valence active*. Nous préférons garder ces termes pour parler de la partition de la valence selon le point de vue sémantique.

Deuxièmement, nous souhaitons utiliser l'analyse distributionnelle pour caractériser la tête d'une unité. Nous cherchons donc à repérer dans le contexte syntaxique d'une unité le contexte qui caractérise plus particulièrement la tête, c'est-à-dire la valence supérieure.

Illustrons ce dernier point par un exemple. Considérons l'énoncé *Une très vieille dame habite ici* et l'unité syntaxique *vieille dame* qu'il contient. Aucun des deux mots *vieille* et *dame* ne contrôlent à lui seul l'entièreté de la distribution de l'unité qu'ils forment ensemble, car *vieille* est connecté à *très* (*très vieille*) et *dame* est connecté à *une* et *habite* (*une dame habite ici*). Bien sûr, une fois que l'on sait que *habite* est la tête de la phrase, on en déduit que la tête de *vielle dame* est *dame*, puisque c'est le mot qui est connecté au gouverneur de cette unité syntaxique.

On ne dira jamais assez qu'on ne peut déterminer la tête d'une unité syntaxique U par la distribution qu'une fois qu'on a déterminé la valence supérieure de U, c'est-à-dire qu'on a déterminé qui parmi les éléments auxquels U est connectée gouverne U. Considérer la distribution sans avoir au préalable distingué valence inférieure et supérieure peut conduire à des erreurs méthodologiques. Il y a là un cercle vicieux dans lequel il faut éviter de rentrer, puisqu'on ne peut appliquer de critère distributionnel pour déterminer la tête d'une unité sans savoir où se trouve son gouverneur. Autrement dit, il s'agit d'un critère pour

déterminer la hiérarchie qui ne peut s'appliquer sans connaître déjà une partie de la hiérarchie.

Supposons que l'on veuille déterminer qui du nom ou du déterminant est la tête du groupe qu'ils forment ensemble (comme par exemple *chat* et *le* dans *le chat dort*). Si l'on décrète que pouvoir être sujet d'un verbe fait partie de la valence supérieure du nom, le problème est biaisé, puisque, en un sens, on a déjà décidé que le nom est la tête du groupe qu'il forme avec le déterminant (on trouve cette erreur dans l'ouvrage de référence d'Igor Mel'čuk de 1988). Tout le problème ici est qu'on ne peut facilement déterminer la valence supérieure du nom seul, car, dans de nombreuses positions, il est indissociable d'un déterminant. (Voir la section 9.24 sur *Nom ou déterminant comme tête*?) Nous avons contourné ce problème en définissant au préalable la structure de connexion (Chapitre 8), puis en hiérarchisant sommairement cette structure en déterminant la tête d'un énoncé (section 9.9), ce qui nous permet de localiser la valence supérieure d'un certain nombre d'unités (mais pas des noms malheureusement!).

9.12 Le critère distributionnel avec effacement

Les critères distributionnels s'appliquent à une unité syntaxique U dont on connaît la valence supérieure, c'est-à-dire le paradigme des gouverneurs possibles (voir section 9.11).

Une tête distributionnelle de U est un élément qui contrôle la valence supérieure de U. Autrement dit, si B dépend de A, alors B ne modifie pas la valence supérieure de A.

Nous allons voir deux critères distributionnels. Le premier, le critère distributionnel avec effacement suppose que A et B ne sont pas mutuellement indissociables pour être appliqué.

Critère distributionnel avec effacement : Si A est autonomisable (et

donc B est effaçable) et que A a la **même valence supérieure** que $U = A \oplus B$ alors A est une **tête distributionnelle** de U. Inversement, si A est autonomisable (et B est effaçable), mais que A n'a **pas** la même valence supérieure que $U = A \oplus B$, alors B est une **tête distributionnelle** de U.

Ce critère généralise le critère d'effacement simple. Le critère d'effacement simple s'applique à une unité U à l'intérieur d'un énoncé donné, sans faire varier le gouverneur de U. Ici on considère la distribution de U, c'est-à-dire tous les gouverneurs potentiels de U.

Illustrons le critère par l'exemple de U=une maison écologique avec A=une maison et $B=\acute{e}cologique$. Ici A et B sont effaçables et on peut donc comparer la valence supérieure de A, B et $U=A\oplus B$. Il est facile de voir que une maison et une maison écologique ont la même valence supérieure (ils peuvent par exemple tous les deux être sujet ou objet d'un verbe ou se combiner avec une préposition) et qu'elle est différente de la valence supérieure de écologique qui ne peut apparaître dans aucun de ces contextes. On en déduit que une maison est la tête de U.

Le critère distributionnel avec effacement s'applique dans des cas où le critère d'effacement simple ne s'applique pas. Reprenons l'exemple de $U=\hat{a}$ Marie. Ici Marie n'est pas effaçable, puisque \hat{a} n'est pas autonomisable. Par contre, \hat{a} est effaçable, on peut donc comparer les valences passives de Marie et \hat{a} Marie. Ces deux unités ont des valences passives complètement différentes. Par exemple, Marie peut être sujet (Marie dort) ou complément d'objet (J'ai vu Marie) et la commutation avec \hat{a} Marie est impossible (*A Marie dort; *J'ai vu \hat{a} Marie). On en déduit que \hat{a} est la tête de \hat{a} Marie.

Le même raisonnement est applicable à *Marie dormait*. Ici encore, *Marie* n'est pas effaçable, mais *dormait* l'est. Comme *Marie* et *Marie dormait* n'ont absolument pas la même valence supérieure, on en déduit *dormait* est bien la tête de la proposition *Marie dormait*.

9.13 Tête effaçable **Q**

Q

Il arrive quelquefois que le critère d'effacement simple et le critère distributionnel avec effacement se contredisent. On parle alors de TÊTE EFFAÇABLE. L'exemple le plus clair est celui de la conjonction de subordination *that* en anglais. Voyons un exemple :

- 9 Tête et dépendance : Hiérarchiser la structure syntaxique
- (30) Mary thinks that Peter slept. 'Mary pense que Pierre dormait.'
- (31) Mary thinks Peter slept. 'Mary pense que Pierre dormait.'

La conjonction that est bien effaçable, ce qui pourrait conduire à considérer que thinks est directement connecté à slept. Mais en même temps on ne peut pas considérer que that dépend de slept car that Peter slept n'a pas la même valence supérieure que Peter slept: si les deux peuvent dépendre de Mary thinks, seul Peter slept peut former un énoncé déclaratif ou bien se combiner avec une conjonction comme if 'sis' ou whether 'si (oui ou non)'. Comme un dépendant ne peut modifier la valence supérieure de son gouverneur, il est donc clair que that n'est pas un dépendant du verbe slept. (Voir l'encadré ?? sur la Co-occupation pour une autre analyse de that.)

(32) Structure de connexion Arbre de dépendance

De tels cas sont rares et le critère d'effacement simple reste un critère simple généralement fiable.

9.14 Le critère distributionnel sans effacement

Le critère distributionnel avec effacement n'est applicable à une unité $U = A \oplus B$ que si A ou B est effaçable. Il ne peut donc être appliqué pour des unités dont les composantes sont mutuellement indissociables comme le radical et sa flexion dans U = chant-ons. On peut utiliser alors un autre critère distributionnel, applicable dans tous les cas.

Critère distributionnelle de U que la commutation de A avec A' peut donner une nouvelle unité U' dont la **valence supérieure** est **différente** de celle de U. Autrement dit, si $U = A \oplus B$, A commute avec A' et U' = A' \oplus B a une valence supérieure différente de A, alors A est une tête distributionnelle de U.

(Voir l'exercice 4, section ??, et sa correction, section ??, pour une autre formulation du critère due à Paul Garde, 1977 : 8.) Si l'on applique ce critère à une

unité $U = A \oplus B$ dont on veut déterminer la tête, on considèrera le paradigme des éléments A' qui commutent avec A et le paradigme des éléments B' qui commute avec B et on comparera les valences supérieures de $U = A \oplus B$, $U' = A' \oplus B$ et $U'' = A \oplus B'$, c'est-à-dire le paradigme des unités X qui peuvent gouverner ces différentes unités.

Figure 9.7 - Application du critère distributionnel sans effacement

On regardera ainsi laquelle des deux commutations, celle sur A ou celle sur B, a le plus d'impact sur la valence supérieure de U. Si les $A \oplus B$ ' ont la même valence supérieure que $A \oplus B$, mais pas les $A' \oplus B$, on en déduira que A gouverne B, c'est-à-dire que A est la tête de la combinaison $U = A \oplus B$.

Montrons par un exemple comment s'applique le critère. Considérons U = chantons qui est la combinaison du lexème A = CHANTER de signifiant chant- et de la flexion verbale $B = -ons = indicatif \oplus présent \oplus 1^{re}$ personne \oplus pluriel (voir le Chapitre ?? sur les *Catégories flexionnels* pour les détails de cette décomposition et la discussion sur la présence de syntaxèmes zéros). On peut faire commuter A avec les autres lexèmes verbaux et B avec les autres flexions verbales. Quel que soit le lexème verbal A' considéré, l'unité U' = A' ⊕ B garde la même valence supérieure que U, celle d'une proposition à l'indicatif, qui peut être aussi bien une phrase complète qu'une proposition subordonnée à un verbe qui impose l'indicatif (Je sais que nous chantons). Par contre, ni U, ni U' ne pourra être subordonnée à un verbe qui demande le subjonctif (*Je veux que nous chantons vs Je veux que nous chantions). Autrement dit, la commutation de B avec un subjonctif changera la valence supérieure de U. Même le changement du présent en futur change la valence supérieure, puisqu'il existe des contextes où le futur est impossible (Si nous chantons, les gens vont partir vs *Si nous chanterons, les gens vont partir). La commutation de B avec un infinitif (chanter) ou avec un participe (chanté ou chantant) modifie encore davantage la valence supérieure de la forme verbale. Nous en concluons que, sans aucun doute possible, c'est la flexion du verbe et non le verbe lui-même qui contrôle la valence supérieure d'une forme verbale.

Le critère distributionnel sans effacement peut également s'appliquer quand le critère distributionnel avec effacement s'applique. Le critère sans effacement est néanmoins moins simple à utiliser que celui avec effacement. En effet, pour ce dernier, il n'y a que trois unités à tester (U = A \oplus B, A et B) et non toutes les unités obtenues en faisant varier A ou B. On peut considérer le critère distributionnel avec effacement comme un cas limite du critère distributionnel sans effacement, une commutation par le vide.

9.15 Tête faible

Reprenons l'exemple de $U = \grave{a}$ Marie, avec $A = \grave{a}$ et B = Marie. Le critère distributionnel avec effacement (section 9.12) nous a permis de déterminer que $A = \grave{a}$ est une tête distributionnelle. Voyons maintenant ce que nous dit le critère distributionnel sans effacement sur les propriétés de tête de A et B. Le syntagme U peut dépendre de téléphoner (*Pierre téléphone à Marie*), mais si on commute $A = \grave{a}$ avec une autre préposition (de Marie, sur Marie), le syntagme peut continuer à dépendre d'un verbe (*Pierre rêve de Marie, Pierre compte sur Marie*), mais pas du verbe téléphoner. On en déduit à nouveau que la préposition \grave{a} est une tête distributionnelle et qu'elle ne dépend donc pas du substantif B. (Pour l'utilisation du terme substantif, voir la section 9.24 Nom ou déterminant comme tête?.) Avant d'en conclure que le substantif B dépend de la préposition A, il faut regarder ce qui se passe quand on commute le substantif B.

Le substantif B = Marie peut commuter avec beaucoup de groupes substantivaux sans que cela n'affecte la valence supérieure du groupe (Pierre téléphone à une amie/à tout le monde/aux autres/...), mais B peut difficilement commuter avec un inanimé (Pierre téléphone à quelqu'un; ^{?*} Pierre téléphone à quelque chose). De plus, B peut commuter avec un infinitif, mais pas dans ce contexte (Pierre commence à partir vs *Pierre téléphone à partir). On en déduit que la préposition à est plutôt la tête distributionnelle, mais que le substantif contrôle quand même une partie de la distribution.

Lorsqu'un deuxième élément contrôle également la distribution de l'unité U, la tête de U est dite faible.

Le groupe prépositionnel est donc une unité à tête faible. Ceci nous permet de postuler la structure de dépendance ci-dessous (figure de gauche). Mais on peut aussi vouloir privilégier le rôle de tête de la préposition et obtenir un arbre de dépendance (figure de droite).

Structure de dépendance Arbre de dépendance avec tête faible

Figure 9.8

Notons qu'il existe une solution plus « lexicale » au fait que la préposition \dot{a}



puisse avoir aussi bien un complément substantif (à Marie) qu'infinitif (à arriver). La solution consiste à considérer que la préposition à a deux entrées lexicales différentes, l'une dont le complément est un substantif et l'autre un infinitif. Dans ce cas, on considère que les verbes TELÉPHONER (téléphoner à Marie) et COMMEN-CER (commencer à arriver) régissent des prépositions à différentes et il n'est plus nécessaire de dire que le complément de la préposition contrôle la valence supérieure du groupe prépositionnel.

9.16 Quelle structure pour quels critères?



En définissant la structure de connexion à partir de critères caractérisant les unités autonomisables (illocutoirement et/ou prosodiquement), nous obtenons une structure qui encode ces mêmes unités (voir la section 8.19 Structure de connexion et unités). Par contre, si l'on ajoute des critères pour raffiner la structure celle-ci n'encode plus les unités autonomisables, comme nous l'avons déjà mentionné dans l'encadré ?? sur Les limites de la dualité. C'est ce qui se passe avec les critères introduits dans ce chapitre qui nous ont permis d'obtenir une structure de dépendance plus fine que la structure de connexion. On peut néanmoins récupérer l'information sur les unités autonomisables en indiquant quelles dépendances ne peuvent être coupées, c'est-à-dire en indiquant pour chaque dépendant s'il est optionnel ou obligatoire.

On peut maintenant se demander quel est l'intérêt d'introduire de nouveaux critères et de raffiner ainsi la structure. Il ne faut pas perdre de vue les objectifs qui prévalent à la définition d'une structure syntaxique. L'un des principaux objectifs de la structure syntaxique est de servir de support à l'écriture des règles de la grammaire. Nous devons donc privilégier la structure qui permet d'exprimer le plus simplement possible les principales contraintes qui pèsent sur la formation des énoncés. C'est pour cela que nous souhaitons que les éléments qui se contraignent les uns les autres soient liés dans la structure. Ceci nous permettra d'assurer les contraintes distributionnelles par des RÈGLES LOCALES, c'est-à-dire des règles qui s'appliquent à une portion bien circonscrite de la structure.

Ajoutons encore que la structure que nous associons à un énoncé représente seulement les propriétés de notre énoncé vis-à-vis des critères appliqués. Nous aimerions que la structure que nous obtenons ait une réalité cognitive, qu'elle soit construite par le locuteur lorsqu'il produit ou comprend un énoncé, mais nous n'avons pas vraiment les moyens de le montrer (voir néanmoins notre tentative dans le Chapitre 2 Produire un énoncé).

9.17 Tête interne et critère rectionnel

La tête distributionnelle est aussi appelée la TÊTE EXTERNE, parce que c'est l'élément qui est « visible » de l'extérieur d'une unité syntaxique. On peut opposer aux critères externes de l'effacement et de la distribution, des critères internes, que nous appellerons les CRITÈRES DE RECTION. Ceux-ci s'appliquent à la combinaison de deux éléments, indépendamment des contextes où peut être placée l'unité formée par leur combinaison et donc sans regard pour ce qui se passe à l'extérieur de cette combinaison.

On appelle TÊTE INTERNE d'une combinaison l'élément qui semble imposer à l'autre sa forme, sa catégorie ou sa position linéaire. La tête interne est encore appelée le RÉGISSANT et l'élément qui dépend de lui est dit RÉGI.

Par exemple, le verbe TÉLÉPHONER impose à son complément d'être combiné avec la proposition À (*Pierre téléphone à Marie*) tandis que le verbe REGARDER demande un complément nu (*Pierre regarde Marie*).

La contrainte que le régissant impose au régi et son expression (par un syntaxème de cas, une préposition ou une place linéaire par exemple) est appelée le RÉGIME. À noter que l'on parle aussi bien du régime du régissant, pour désigner l'ensemble des contraintes qu'il impose à ses dépendants, que du régime du régi, pour désigner la forme particulière que lui impose son gouverneur.

Une préposition imposée par un régime, comme la préposition À imposée par le verbe TÉLÉPHONER à son complément, est appelée une PRÉPOSITION RÉGIME.

Le critère rectionnel permet par exemple de confirmer que l'auxiliaire est bien la tête d'une forme verbale complexe. Comparons :

(33) Pierre a dormi longtemps.

(34) Pierre va dormir longtemps.

On voit que l'auxiliaire du passé impose au verbe lexical d'être au participe passé, tandis que l'auxiliaire du futur impose au verbe d'être à l'infinitif. On en déduit que l'auxiliaire régit le verbe auxilié.

Une formulation inverse du critère rectionnel est le Critère d'indépendance des codépendants. On appelle codépendants deux éléments qui dépendent du même regissant.

Un élément n'impose pas à un codépendant sa forme, sa catégorie ou sa position linéaire.

Un élément peut toutefois occuper la place habituellement dévolue à un de ses codépendant, mais il ne contrôle pas la place où celui-ci est déplacé.

Donnons l'exemple du wolof. Le verbe en wolof est quasiment invariable, mais il peut se combiner avec un certain nombre de particules verbales. La question est alors : ces particules doivent-elles être traitées comme des adverbes modifiant le verbe ou comme des auxiliaires régissant le verbe ? Ces particules verbales modifient la distribution du verbe (par exemple, les particules ne sont pas possibles dans les constructions relatives), ce qui en fait donc des têtes distributionnelles. Mais elles ont une autre propriété qui en fait aussi des têtes internes. En effet, chaque particule promeut un élément différent en première position : la particule *a* maintient le sujet en première position, la particule *la* promeut un complément, la particule *na* promeut le verbe lui-même et enfin la particule *da* occupe la première position.

- (35) ñu=a=y lekk am xar S3PL=PART=IMP manger INDEF mouton 'C'est eux qui mangent un mouton.'
- (36) am xar $la=\tilde{n}u=y$ lekk INDEF mouton PART=S3PL=IMP manger 'C'est un mouton qu'ils mangent.'

- 9 Tête et dépendance : Hiérarchiser la structure syntaxique
- (37) lekk **na**=*ñu*=*y* am xar manger PART=S3PL=IMP IÀÀNDEF mouton 'Ils mangent un mouton.'
- (38) da-ñu=y lekk am xar
 PART=S3PL=IMP manger INDEF mouton
 '(Le fait est que) ils mangent un mouton.'

En plus, la particule *la* autorise la réalisation d'un sujet lexical post-verbal, tandis que les particules *na* et *da* l'interdisent. Ces propriétés montrent que les particules ne peuvent pas être dépendantes du verbe, en raison du critère d'indépendance des codépendants : un dépendant du verbe ne peut imposer sa place à un autre dépendant du verbe comme le font les particules verbales.

Notons pour terminer que lorsque les critères pour identifier la tête interne et la tête externe se contredisent, ce sont les critères distributionnels (avec et sans effacement) et la tête externe que nous privilégions.

9.18 Critère de la tête sémantique

La tête syntaxique d'une unité syntaxique n'est en général pas seulement saillante du point de vue distributionnel : elle est aussi saillante du point de vue sémantique et constitue alors ce qu'on appelle la tête sémantique de l'unité.

La TÊTE SÉMANTIQUE d'une unité U est l'élément de U qui résume à lui seul le sens de U, qui en détermine la catégorie ontologique.

Par exemple, la tête sémantique de *poisson-chat* est *poisson*, car un poisson-chat est un type de poisson et pas un type de chat. À l'inverse, *enseignant-chercheur* apparaît comme exocentrique, car un enseignant-chercheur est autant un enseignant qu'un chercheur. (Voir l'encadré ?? *Constructions* N+N ou $N \oplus N$?.)

La tête sémantique d'une proposition est le verbe. Ainsi la proposition *Zoé invite ses copains à son anniversaire* décrit-elle avant tout une invitation dont Zoé, ses copains ou l'anniversaire sont des participants.

La tête sémantique d'une unité syntaxique n'est pas toujours sa tête syntaxique : la notion de tête sémantique privilégie les sémantèmes lexicaux vis-à-vis des éléments grammaticaux. Ainsi la tête sémantique de *Zoé a invité ses copains à son anniversaire* reste le verbe lexical inviter et n'est pas l'auxiliaire avoir qui marque le passé. De la même façon, la tête sémantique d'un groupe substantival comme *le chien de Marie* privilégie le nom tête CHIEN au détriment du déterminant LE.

La notion de tête sémantique est donc une notion distincte de celle de tête syntaxique (voir l'encadré ?? sur les *Distorsions sémantique-syntaxe*). Néanmoins lorsque les critères pour déterminer la tête syntaxique sont insuffisants, on peut, si l'on souhaite à tout prix choisir une tête, utiliser les critères sémantiques.

9.19 Critère sémantique de non-effacement

On ne confondra pas notion de tête sémantique avec celle de GOUVERNEUR SÉMANTIQUE (voir l'encadré ?? sur les Dépendances morphologique, syntaxique et sémantique). Par exemple, un livre rouge a pour tête sémantique livre, car un livre rouge est avant tout un livre et pas un objet rouge. On peut par exemple s'en convaincre avec :

(39) Le livre rouge est plus beau que les autres.

En effet, ici clairement, *les autres* renvoie aux autres livres et pas aux autres objets rouges. (On notera au passage que cette asymétrie disqualifie les formalisations logiques symétriques qui voient *un livre rouge* comme un *x* tel que *x* est un livre et *x* est rouge.) Pourtant, dans la relation sémantique entre 'livre' et 'rouge', c'est bien 'rouge' qui fonctionne comme prédicat et prend 'livre' comme argument ('rouge' désigne une propriété qui comme toute propriété est la propriété de quelque chose, alors que 'livre' désigne un objet dont la couleur n'est pas définitoire). On dira donc que *livre* est la tête sémantique de *livre rouge*, tandis que 'rouge' est le gouverneur sémantique de 'livre', qui est son argument ou dépendant sémantique.

Un élément X comme *rouge* qui prend son gouverneur syntaxique comme argument sémantique, c'est-à-dire qui est à la fois dépendant syntaxique et gouverneur sémantique d'un élément Y est appelé MODIFIEUR de Y (voir le Chapitre ?? sur la *Structure syntaxique profonde*). Les modifieurs sont toujours effaçables a priori car ils ne sont pas indispensables à leur gouverneur syntaxique. Ceci nous donne un nouveau critère sémantique qui est le pendant du critère d'effacement simple (il s'applique quand le critère d'effacement simple ne s'applique pas) et que nous appelons donc le critère sémantique de non-effacement.

Critère sémantique de non-effacement : si l'unité U = AB est gouverné syntaxiquement par X, que A est le **gouverneur sémantique** de B et A est **non effaçable**, alors A est une **tête syntaxique** de U.

Autrement dit, si A est non effaçable, il ne peut être un modifieur de B. Comme il est le gouverneur sémantique de B, il ne peut être son dépendant syntaxique et il a donc des propriétés de tête.

Prenons un exemple. On s'intéresse à U = sur la table dans l'énoncé suivant :

(40) Pierre a posé le livre sur la table.

Ici 'sur' est un prédicat binaire qui exprime une relation de localisation entre 'livre' et 'table' et prend donc 'table' comme argument. Comme A = sur est le gouverneur sémantique de $B = la \ table$ et que A est non effaçable, A est une tête syntaxique de B.

D'après le critère d'effacement simple, si B est effaçable, alors A est une tête syntaxique, indépendamment de savoir si A gouverne sémantiquement B ou pas. Le critère du gouverneur sémantique est donc vraiment intéressant quand ni A, ni B n'est effaçable.

On peut encore appliquer le critère au cas des formes verbales comme *chant-ait*. C'est la flexion qui gouverne sémantiquement le lexème verbal : le temps verbal exprime un sens fondamentalement prédicatif ('avoir lieu dans le passé') dont le signifié du lexème verbal est l'argument. Nous verrons une autre application pour le cas de la connexion déterminant-nom à la section 9.27 *Déterminant comme tête*?

9.20 Distorsions sémantique-syntaxe

Indépendamment de la question des éléments grammaticaux, il existe certaines configurations de deux éléments lexicaux où tête syntaxique et tête sémantique ne semblent pas se correspondre. On parle alors de distorsions (en anglais de MISMATCHES). Le plus flagrant des exemples de ce type est mis en évidence par la paire minimale suivante, déjà utilisée à de nombreuses reprises par des linguistes (le Dictionnaire de l'Académie Française de 1798 contient déjà deux entrées séparées pour VERRE):

- (41) Félix achète un verre à vin (et le casse).
- (42)Félix achète un verre de vin (et le boit).

On peut supposer que ces deux exemples ont la même structure syntaxique, que dans les deux cas un verre est la tête syntaxique de l'objet de achète et que vin est complément de la préposition à ou de et dépend de un verre par l'intermédiaire de cette préposition. Par contre, du point de vue sémantique, les situations sont très différentes : si ce que Félix casse est bien un verre, ce qu'il boit est du vin. Dans un verre de vin, un verre sert uniquement à mesurer la quantité de vin et pas à déterminer la catégorie ontologique de la chose désignée par le syntagme un verre de vin. On en déduit donc que la tête sémantique de verre à vin est verre, tandis que la tête sémantique de verre de vin est vin. Il y a ainsi une distorsion sémantique-syntaxe dans la phrase Félix achète/boit un verre de vin. Nous représentons ci-dessous son arbre de dépendance syntaxique et son graphe sémantique (pour ce dernier on consultera les chapitres 2, ?? ou ??):

(43)

Un autre cas de distorsion est illustré par l'exemple suivant :

Il paraît que le président était au courant.

Nous savons que la tête syntaxique de cette phrase est la forme verbale *paraît*. Pourtant, à une telle déclaration, un interlocuteur peut répondre « Non, je ne crois pas. », ce qui en général voudra dire 'je ne crois pas que le président était au courant' et non pas 'je ne crois pas qu'il paraît que le président est au courant'. Autrement dit, tout se passe comme si il paraît que était un modifieur de le président était au courant. Cet emploi du verbe PARAÎTRE a déjà été qualifié de RECTEUR FAIBLE pour cette raison par Claire Blanche-Benveniste1989. Ce verbe s'emploie d'ailleurs dans une autre construction où non seulement il n'est pas la tête sémantique, mais il n'est plus non plus la tête syntaxique, la relation de rection n'étant plus exprimée :

Le président était au courant, il paraît.

On notera également que dans d'autres langues, notamment une langue agglutinante comme le japonais, un sens comme 'il paraît' est réalisé par un affixe verbal. Ainsi, le verbe XXX tsūjiru 'comprendre' est suivi d'une série d'affixes modaux terminant sur ⊠⊠ sōda 'il paraît'.

(46) XXXXXXXXXXXXXXXXXX

Gichō-wa jijō-ni tsūjite ita **sōda**. président-topique circonstances-datif saisir-en-train être.passé **paraît**. 'Il **paraît** que président était au courant.'

Donnons un dernier exemple qui montre que le choix de la tête syntaxique dépend parfois davantage de la façon dont le lexique de la langue est organisé que d'un choix conscient du locuteur. Voici un exemple classique de distorsion lors d'une traduction :

(47) Antoine traversa la rivière à la nage. *Antony swam accross the river*. lit. Antony nagea à travers la rivière

Cette **réorganisation syntaxique** a été décrite par **Tesnière1959** : Chapitre ??) et nommé par lui MÉTATAXE. Le phénomène a été rendu fameux par les études de Leonard Talmy (voir son article de 1983, How language structures space, ou son ouvrage de 2003, Toward a cognitive semantics). Talmy montre que certaines langues, comme le français, favorise le déplacement (path-incorporating), lequel est exprimé par le verbe (entrer, sortir, monter, descendre, traverser, etc.), alors que celui-ci est exprimé par une préposition ou une particule dépendant du verbe en anglais (in, out, up, down, accross, etc.). D'autres langues, comme l'anglais, favorise à l'inverse la manière (manner-incorporating) dont le mouvement a lieu (swim 'nager', run 'courir', drive 'se déplacer en voiture', bike 'se déplacer en vélo', etc.), alors que celle-ci est généralement exprimée par un modifieur en français (à la nage, en courant, en voiture, en vélo, etc.). Il ne s'agit pas à proprement parler d'un exemple de distorsion syntaxe-sémantique (puisque finalement chaque langue adapte sa hiérarchie sémantique et syntaxique à son lexique dans ce cas), mais plutôt d'une illustration du fait que la hiérarchie syntaxique ne découle pas seulement d'une hiérarchie de l'information à communiquer, mais aussi de propriétés de la langue qui sert à l'exprimer.

9.21 Synthèse des critères pour la tête syntaxique

Bien qu'il nous reste encore un critère à présenter (le critère de recouvrement), le moment nous semble venu de faire une synthèse des différents critères pour déterminer la tête d'une unité syntaxique. Nous considérons neuf critères, quatre purement syntaxiques et quatre mettant en jeu d'autres niveaux d'analyse.

Les **quatre critères purement syntaxiques** pour déterminer la tête syntaxique d'une unité U sont par ordre d'importance :

- le critère distributionnel avec effacement : un élément autonomisable de U ne peut être la tête syntaxique de U que s'il a la même valence supérieure que U (section 9.12).
- le critère distributionnel sans effacement : la tête syntaxique de U est l'élément qui contrôle la valence supérieure de U (3.3.13) et porte en conséquence les marques caractéristiques de la position syntaxique occupée par U; la tête syntaxique d'un énoncé U est l'élément qui porte les marques caractéristiques de l'illocution (section 9.9);
- le critère d'effacement simple : la tête syntaxique de U est l'élément qui peut le plus facilement former une unité syntaxique avec le gouverneur de U, ce qui revient à effacer les autres éléments de U (section 9.8).
- le critère rectionnel : la tête syntaxique de U est l'élément qui peut imposer aux autres éléments de U leur forme, leur catégorie ou leur place linéaire (section 9.17).

Les trois autres critères concernent les relations avec les autres niveaux d'analyse. Il repose sur l'hypothèse qu'il existe une certaine similitude entre les différents niveaux d'organisation de l'énoncé et que donc si les critères purement syntaxiques ne sont pas déterminants, autant maximiser cette similitude. Voici donc, par ordre d'importance, **cinq critères additionnels** pour déterminer la tête syntaxique d'une unité U :

- le critère sémantique de non-effacement : un élément non effaçable qui gouverne sémantiquement les autres composantes de U est une tête sémantique (section 9.19)
- le critère topologique (critère de recouvrement): la tête syntaxique de U
 n'est généralement pas recouverte par une connexion entre deux éléments de U (section 9.33);
- le critère prosodique : la tête syntaxique de U est l'élément qui peut le plus facilement former une unité prosodique avec le gouverneur de U (section 9.8);
- le critère morphologique : la tête syntaxique de U tend à contrôler les marques morphologiques du gouverneur de U (section ??); la réciproque le fait que le gouverneur de U contrôle les marques morphologiques de la tête de U est déjà exprimé dans le critère distributionnel sans effacement;
- le critère de la tête sémantique : la tête syntaxique de U est généralement l'élément de U qui résume à lui seul le sens de U (section 9.18).

9.22 Les cas problématiques

Dans l'ensemble, les critères pour l'attribution de la tête d'un syntagme sont assez convergents. Par exemple, on s'accorde en général pour dire que l'objet dépend du verbe sans qu'il n'y ait de réelles discussions sur ce point (à l'exception néanmoins des constructions à verbes supports que nous discutons dans l'encadré ?? Quand le dépendant contrôle son gouverneur). Même pour le sujet, tout le monde semble aujourd'hui d'accord pour dire qu'il dépend de la forme verbale et la discussion porte plutôt sur la granularité de la représentation et la question de savoir si le sujet dépend d'un seul des éléments de la forme verbale (la flexion) ou s'il dépend du tout.

Il existe néanmoins un certain nombre de cas problématiques qui sont au centre des discussions sur la notion de tête. Voici les principaux cas discutés dans ce chapitre ou dans la partie 5 :

- les marqueurs de rection : nous avons déjà évoqué dans ce chapitre la question de la tête de à Marie dans Pierre téléphone à Marie. Le même genre de question se pose pour la conjonction de subordination que dans Pierre pense que Marie dort, d'autant que le verbe régi peut être porteur d'une marque de rection (le subjonctif), comme dans Il faut que Pierre dorme, et qu'il contrôle donc en partie la valence supérieure (voir aussi la discussion sur les Têtes effaçables dans l'encadré ??). Ces questions seront à nouveau traitées dans le Chapitre ?? lorsque nous discuterons la translation.
- la coordination : quelle est la tête de Marie et Pierre dans J'ai invité Marie et Pierre? À peu près toutes les configurations ont été envisagées : 1) la construction est symétrique et Marie et Pierre sont des co-têtes; 2) la construction est asymétrique et le premier conjoint, Marie, est l'unique tête;
 3) la conjonction de coordination est l'unique tête. Cette question sera étudiée dans un chapitre entièrement consacré à la coordination et aux autres constructions de ce type, le Chapitre ?? sur les Entassements.
- la relative : quelle est la tête de la relative à qui tu parlais dans la phrase La fille à qui tu parlais habite dans ma rue? S'agit-il du pronom relatif ou bien du verbe principal de la relative? Cette question sera discutée en détail au Chapitre ?? consacré à l'Extraction, dont la relativisation est un cas particulier.
- le déterminant : quelle est la tête de le chien dans Le chien dort? Le déterminant ou le nom? Nous aborderons cette question très controversée dans les prochaines sections.

9.23 « Déterminants complexes » 🖟

Le français possède des syntagmes de la forme « N de N » ou « Adv de N », un tas de gens ou trop de gens, où le nom qui suit de est clairement la tête sémantique : un tas de gens désigne des gens, pas un tas. De plus, dans ces constructions, l'accord se fait généralement avec le nom qui suit de, ce qui, en vertu du critère morphologique, désigne également ce nom comme tête :

- (48) Un tas de lettres ont été écrites depuis le front pendant la guerre de 14.
- (49) La plupart de ces gens sont d'accord avec moi.

Une première analyse consiste à traiter le segment qui précède le nom comme un DÉTERMINANT COMPLEXE en raison de ces propriétés et de la commutation apparente avec un déterminant, comme le montre la commutation possible ici de *un tas de* avec *plusieurs*:

(50) Plusieurs lettres ont été écrites depuis le front pendant la guerre de 14.

Si cette analyse peut être défendue au niveau syntaxique profond (voir Chapitre ??), elle est difficile à défendre au niveau syntaxique de surface, car la commutation n'est plus possible dés qu'il y a une coordination :

- (51) Un tas de lettres et de paquets sont arrivés aujourd'hui.
- (52) *Plusieurs lettres et de paquets sont arrivés aujourd'hui.

Si l'analyse de *un tas de* comme déterminant ne peut être retenue, reste la question de savoir qui du nom *tas* ou de *lettres* gouverne l'autre dans *un tas de lettres*. Bien que, comme nous venons de le voir, les critères morphologique et sémantique donnent l'avantage à *lettres*, on peut considérer que c'est bien *tas* qui est la tête syntaxique. Cela tient au fait que *de*, dans *un tas de lettres*, est une préposition régie par *tas* (critère rectionnel). Les arguments pour cela sont que, d'une part, *de* doit être répété en cas de coordination, comme on l'a vu plus haut. Et que, d'autre part, *de lettres* se pronominalise par le pronom *en* (*j'en ai écrit un tas*), comme pour les autres emplois de *de* (*j'en rêve*). Remarquons que le critère distributionnel ne s'applique pas vraiment (nous comparons deux noms, donc deux items de distribution comparable) et le critère d'effacement simple n'est pas non plus concluant, puisque le premier nom peut être effacé (*des lettres ont été écrites depuis le front*; *ces gens sont d'accord avec moi*), comme le second (? *un*

tas ont été écrites depuis le front; la plupart sont d'accord avec moi). (Remarque : nous pouvons considérer que de est un amalgame de + des dans un tas de lettres; cf. il parle de lettres et pas *il parle de des lettres).

Le cas des constructions « Adv de N » est similaire. Ici aussi, de apparaît bien comme une préposition régime, comme le montre la coordination et la pronominalisation :

- (53) J'ai mangé trop **de** chocolat et **de** pain.
- (54) J'en ai mangé trop.

On peut aussi noter que l'adverbe peut régir un complément qui se place au-delà du nom, ce qui est une raison supplémentaire pour supposer qu'il gouverne le nom (voir la section 9.33 sur la *Critère de recouvrement*) :

- (55) Il a fourni **trop** d'efforts **pour abandonner maintenant**.
- (56) Il a envoyé plus de lettres que nécessaire.

Ceci conduit donc à considérer l'adverbe comme la tête du syntagme.

(57)

Mais il faut néanmoins noter que ce syntagme n'est pas toujours continu :

(58) J'ai **trop** mangé **de chocolat**.

Et même pour les adverbes négatifs, dans certains cas, il ne peut pas être continu :

- (59) Je n'ai jamais mangé de chocolat.
- (60) *Je n'ai mangé jamais de chocolat.

Ceci peut amener à se demander si dans cette dernière construction, *de chocolat* dépend bien de l'adverbe ou s'il ne dépend pas du verbe, ce qui est l'analyse traditionnelle, *de* étant alors déclaré déterminant négatif (voir la discussion sur les différents emplois de *de* dans l'encadré ?? sur *Les articles indéfinis du français*). Cet argument n'est plus tenable en l'absence d'un verbe :

(61) Merci, jamais de chocolat pour moi la semaine. Trop de chocolat le week-end.

9.24 Nom ou déterminant comme tête?

Le déterminant et le nom forment avec leurs dépendants une unité que nous appelons le GROUPE SUBSTANTIVAL. Le groupe substantival est souvent appelé GROUPE NOMINAL OU GROUPE DÉTERMINATIF, selon que l'on considère que sa tête est le nom ou le déterminant. Nous avons choisi à dessein un terme neutre qui ne présuppose aucune des deux analyses. Qui plus est ce terme est conforme aux choix terminologiques fait en typologie qui distingue généralement SUBSTANTIF (une unité qui peut référer à une substance) et NOM (une catégorie lexicale). (Pour plus de détails, voir la section ?? sur Substantifs et noms.)

La question de savoir qui du nom ou du déterminant est la tête est une question complexe qui a fait couler beaucoup d'encre. Il existe trois réponses possibles (et pas deux) : (A) le nom est la tête, (B) le déterminant est la tête, (C) les deux ont des propriétés de têtes et sont des co-têtes. Illustrons ces trois options sur un exemple simple, *le petit chien dort* :

(A) Nom tête (B) Déterminant tête (C) Co-têtes

Figure 9.9

L'option (A) est celle de la grammaire traditionnelle, qui utilise le terme groupe nominal. Elle a été sérieusement remise en question dans les années 1980, notamment par Richard Hudson1984 et la thèse de Steven Abney1987 qui défendait l'hypothèse du groupe déterminatif (*DP-hypothesis*), et par l'adoption ensuite du DP (*Determiner Phrase*) par la syntaxe X-barre en lieu et place du NP (*Noun Phrase*). La troisième option est sérieusement envisagée par Lucien Tesnière1959, qui considère que les articles sont des indices qui marquent le rôle nominal du nom, et qui donc traite le chien comme un nucléus translatif, lequel est représenté horizontalement, comme dans notre figure (C) (voir l'encadré ?? sur l'Historique des représentations syntaxiques par des diagrammes en dépendance).

Nous allons discuter les arguments pour les différentes options.

9.25 Nom comme tête?

Le choix du nom comme tête du groupe substantival (option (A) de la section précédente) prend comme argument principal que, en l'absence de critères syntaxiques déterminants, c'est la tête sémantique qui doit primer, laquelle est le nom. De plus, des restrictions de sélection s'exercent sur le nom, puisque certains verbes par exemple ne peuvent prendre que des noms humains comme

9 Tête et dépendance : Hiérarchiser la structure syntaxique

argument:

- (62) Personne ne pense ça.
- (63) ??Rien ne pense ça.

Le cas le plus net de contrôle de la valence supérieure par le nom est celui des noms temporels, qui peuvent modifier un verbe :

- (64) Elle est venue trois jours / quelques minutes / un samedi matin.
- (65) *Elle est venue trois tables / quelques amis / une cuillère.

Cette opposition montre bien que la valence supérieure de *trois jours* est déterminée par *jour* et non par *trois* (Van Langendonck1994) (critère distributionnel sans effacement). Le même argument vaut pour des noms comme *fois* qui ont la valence supérieure d'un adverbe tout en se combinant avec toute sorte de déterminants comme un nom ordinaire :

(66) Il a bu **trois fois plus** de vin que les autres.

Par ailleurs, un certain nombre de noms s'utilisent sans déterminant (*lundi, maman, avril, Paris, Sylvain*, etc.) et peuvent recevoir un déterminant sans que cela change leur valence supérieure (critère distributionnel avec effacement): *le lundi, ma maman, le Paris de mon enfance, le Sylvain que je connaissais*, etc. (Notons quand même que seul le nom nu peut être utilisé en vocatif, pour interpeler quelqu'un par exemple : *Maman, aide-moi*!)

Un des arguments les plus utilisés pour le déterminant comme tête est l'unicité du déterminant. Il est vrai que le déterminant est un élément particulier du groupe substantival, par sa position initiale notamment. Néanmoins, son unicité reste discutable. D'abord, beaucoup de langues n'ont pas de déterminants (notamment les langues slaves, qui sont les langues maternelles de beaucoup de linguistes qui ont influencé le développement de la syntaxe de dépendance) et dans ces langues le nom est donc incontestablement la tête du groupe substantival. Similairement, les langues germaniques, qui possèdent bien des déterminants, utilisent le nom nu à l'indéfini lorsqu'il est massif ou pluriel (autrement dit, il n'y a pas de morphème traduisant les articles français du et des):

(67) He drinks milk / girly cocktails.

'Il boit du lait / des cocktails de midinettes.'



Ces configurations pourraient s'analyser à l'aide d'un déterminant zéro (c'est-àdire sans réalisation phonologique), mais elles militent quand même en faveur d'une analyse en groupe nominal, beaucoup plus simple. En effet si le déterminant est effaçable sans que la valence supérieure ne change, il n'est pas la tête distributionnelle (critère distributionnel avec effacement).

Le français présente lui aussi des données problématiques pour l'unicité du déterminant:

- (68)mes deux amis
- (69)mes amis
- (70)deux amis

Doit-on considérer que dans deux amis, deux est déterminant? Mais alors, dans mes deux amis, y a-t-il deux déterminants? Ou bien s'agit-il d'un autre deux? Cette dernière solution est la solution la plus souvent proposée, mais elle est difficilement défendable étant donné que deux a absolument la même position linéaire et la même contribution sémantique que les soit ou non présent et qu'il en va de même de tous les éléments qui peuvent commuter avec lui, comme quelques ou différents (voir l'exercice 5 du Chapitre 11). Une application stricte du critère distributionnel avec effacement nous oblige à conclure que mes deux amis et deux amis ont la même valence supérieure et que donc mes n'est pas la tête distributionnelle. Voir l'encadré ?? sur les Grammèmes comme tête pour les différentes structures syntaxiques possibles de mes deux amis.

9.26 Quand le dépendant contrôle son gouverneur 🍳



Il existe d'autres arguments pour le nom comme tête que nous n'avons pas retenus. Parmi ceux-ci, il est souvent avancé que le nom contrôle son déterminant. On dit A CONTRÔLE B lorsque A réduit le paradigme des B par rapport à ce qui serait attendu.

Les noms massifs prennent comme article indéfini du et non un : du sable vs. #un sable (voir le Chapitre ?? pour une description détaillée des articles du français). De plus, ils sont non comptables et excluent donc tous les déterminants quantifieurs: #deux sables, #plusieurs sables (ces expressions sont possibles, mais sables est alors interprété comme désignant des 'types de sable').

Cela ne permet pas d'en déduire que le nom est gouverneur, car il est tout à fait possible qu'un dépendant contrôle son gouverneur. Ceci est assez courant dans

les **collocations** (voir la section 6.11 Collocation et choix lié). Dans les constructions à verbe support (**pousser** un cri, **mener** une lutte, **donner** un avis, **poser** une question, etc.), un nom prédicatif **contrôle lexicalement** le choix du verbe, alors qu'il en est le complément d'objet. Dans les constructions locatives (aller **sur** la plage, **dans** la montagne, à la campagne, **en** Chine, **chez** le docteur, etc.), le nom sélectionne la préposition locative, alors qu'il en est le complément. Le dépendant peut aussi contrôler son gouverneur dans les **phénomènes d'accord** : ainsi le sujet contrôle-t-il l'accord du verbe qui le gouverne.

9.27 Déterminant comme tête?

Le choix du déterminant comme tête du groupe substantival (option (B) de la section 9.24 Nom ou déterminant comme tête?) repose sur différents arguments. D'abord, une analogie entre la proposition et le groupe substantival : si l'on considère que l'auxiliaire, qui portent les grammèmes du verbe, est la tête de la proposition, on peut être amené à considérer que l'article, qui a aussi un rôle essentiellement grammatical, est la tête du groupe substantival (les théories issues du générativisme parle alors de têtes fonctionnelles pour désigner ces éléments et les opposent aux têtes lexicales). On notera qu'en français, en plus d'être le porteur de la définitude, c'est phonologiquement l'article qui porte le nombre et non le nom (le chat /ləfa/ vs. les chats /lefa/) (pour plus de détail voir la discussion dans le Chapitre ?? sur les Catégories flexionnelles). En allemand, l'article porte plus régulièrement le cas que le nom et contrôle ainsi la valence supérieure du groupe substantival (der Mann, 'l'homme' au nominatif, sera sujet, tandis que den Mann, à l'accusatif, sera objet). En français aussi, l'article est en quelque sorte porteur du cas, si l'on considère les amalgames entre les prépositions \hat{a} et de, marqueurs de cas, et les articles comme des formes casuelles de l'article : le lit vs. (aller) au lit vs. (sortir) du lit; les amis vs. (parler) aux amis vs. (parler) des amis; des jouets vs. (parler) de jouets. En appliquant le critère distributionnel sans effacement, on en déduit que l'article est la tête, car la commutation de le par au ou du (ou de der par den en allemand) change la valence supérieure.

Un deuxième argument exclut de traiter l'article comme un dépendant ordinaire du nom. Rappelons qu'un dépendant ne doit pas changer la valence supérieure de son gouverneur. Or il existe des positions syntaxiques où le déterminant est obligatoire et d'autres où il est au contraire proscrit.

- (71) La syntaxe est utile.
- (72) *Syntaxe est utile.

- (73) Nous parlons syntaxe.
- (74) *Nous parlons la syntaxe.

On en déduit que *la syntaxe* et *syntaxe* n'ont pas la même valence supérieure (critère distributionnel avec effacement).

Dans le même ordre d'idée, il existe des déterminants qui ne sont possibles que dans certains contextes, comme *de* ou *ni* qui s'utilisent uniquement dans un contexte négatif :

- (75) Bernd ne mangeait pas de viande.
- (76) *Bernd mangeait de viande.
- (77) Bernd ne mangeait ni viande, ni poisson.
- (78) *Bernd mangeait ni viande.

Une autre analyse, est de considérer que *de* n'est pas un déterminant, mais le régime de *pas* (voir section ??), et que *ni* est une conjonction de coordination, mais on est alors ramené au premier point, à savoir des positions où le déterminant est proscrit, ce qui est aussi un argument pour le déterminant comme tête (critère distributionnel avec effacement).

Il existe encore d'autres contextes où seul le nom nu est possible, comme après la préposition *sauf* en début de phrase :

- (79) Sauf abonnement souscrit pour une durée déterminée [...], le client peut interrompre son abonnement à tout moment par lettre recommandée avec demande d'accusé *de réception*. (moncompte.lemonde.fr)
- (80) Sauf sursaut d'ici à un prochain « sommet » européen en avril, les historiens dateront certainement de cette affaire, de ces années 2015-2016, le début de la décomposition de l'Europe. (lemonde.fr, à propos de la crise des réfugiés syriens)
- (81) *Sauf l'abonnement ...
- (82) *Sauf un sursaut ...

Un troisième argument est la possibilité pour certains déterminants de pouvoir être utilisés sans le nom, comme *plusieurs*, *certains*, *un*, *deux* ... (critère d'effacement) :

- (83) Plusieurs amis sont restés dormir.
- (84) (Des amis sont venus.) Plusieurs sont restés dormir.

On peut répondre à ça qu'il s'agit dans le deuxième cas d'un pronom de même forme que le déterminant et que, pour d'autres déterminants, le pronom correspondant n'a pas la même forme :

- (85) Chaque ami a payé sa part.
- (86) (Des amis sont venus.) Chacun / *Chaque a payé sa part.

Reste que de manière général, les pronoms, qui forment à eux seuls des groupes substantivaux, sont issus de déterminants, soit qu'ils aient la même forme comme plusieurs, soit une forme morphologiquement dérivée comme chacun (de chaque + un) ou celui (de ce + lui). On peut d'ailleurs retourner cette analyse, comme le suggère Richard Hudson2007, et considérer que ce ne sont pas les déterminants qui ont des emplois de pronoms, mais que c'est les pronoms qui ont des emplois de déterminants. Autrement dit, les déterminants sont des pronoms transitifs (c'est-à-dire prenant un complément nominal nu), certains ayant un complément obligatoire (comme des ou chaque) et d'autres ayant un complément optionnel (comme un ou plusieurs) (critère rectionnel). (Voir l'encadré ?? sur la Co-occupation pour une analyse qui va dans la même direction.)

En allemand, tout déterminant (autre que *ein* 'un' ou *kein* 'aucun') peut s'utiliser sans le nom, c'est-à-dire comme pronom (critère d'effacement) :

- (87) Ich habe **den/diesen Apfel** weggeschmissen. 'J'ai jeté **la/cette pomme**.'
- (88) Ich habe den/diesen weggeschmissen. 'Je l'ai jetée. / J'ai jeté celle-là.'

On peut à nouveau postuler un pronom de forme identique pour chaque déterminant, mais cette analyse est mise à mal par l'existence de « vrais » pronoms comme *ihn*. Les deux types d'éléments ont d'ailleurs des distributions différentes, puisque, plus facilement encore qu'en français, les déterminants s'emploient avec un adjectif et sans nom, ce qui est exclu pour le « vrai » pronom :

- (89) Ich habe **den/diesen faulen** weggeschmissen. 'J'ai jeté **la/celle pourrie**.'
- (90) Ich habe **ihn** (*faulen) weggeschmissen. 'Je l'ai jetée.'

Un autre argument encore concerne les « déterminants complexes » (voir l'encadré ?? sur les « *Déterminants complexes »*) comme *trop de* ou *des tas de* :

(91) Félix a invité **trop d**'amis / **des tas d**'amis.

Ces segments excluent l'article défini (**trop des amis ou *des tas des amis), mais ils acceptent la présence d'un déterminant possessif (trop de mes amis, des tas de mes amis). Ces données montrent que les amis et mes amis n'ont pas la même valence supérieure, ce qui est un argument pour le déterminant comme tête (critère distributionnel sans effacement).

Nous avons vu que les noms propres pouvaient éventuellement prendre un déterminant (le Paris de mon enfance), ce que nous avons utilisé comme argument pour le nom comme tête (section 9.25 Nom comme tête). Une analyse plus fine des données peut conduire à la conclusion inverse. En fait, les noms propres ne peuvent pas être facilement modifiés sans l'ajout d'un déterminant : ?*Paris de mon enfance était magnifique, ?*Sylvain que je connaissais avait beaucoup de cheveux, ?*Maman de Zoé est en retard. On peut confirmer que les substantifs ne sont généralement pas modifiables comme les noms, comme le montre le cas de quelqu'un, qui ne peut pas être modifié par un adjectif sans l'ajout d'un marqueur (on dit quelqu'un de sympa et pas ?? quelqu'un sympa). Le rôle de le dans le Paris de mon enfance est finalement assez paradoxal : alors que son rôle est normalement de permettre à un nom d'occuper une position de substantif (on dit qu'il translate le nom en substantif, voir Chapitre ??), en se combinant à un mot qui est déjà substantif comme Paris, il permet à ce mot d'être rétrogradé au statut de nom et de pouvoir être modifié comme un nom.

Qu'est ce que cela nous dit sur le statut du déterminant comme tête du nom? Le critère distributionnel avec effacement revient à dire que l'ajout d'un dépendant ne doit pas changer la valence supérieure de son gouverneur. Notre argument est ici similaire mais différent : nous disons que l'ajout d'un déterminant sur un nom propre change sa valence *inférieure* et que l'ajout d'un dépendant ne devrait pas non plus changer la valence inférieure de son gouverneur. Et donc *le* dans *le Paris de mon enfance* ne se comporte pas comme devrait se comporter un dépendant.

Un dernier critère joue en faveur du déterminant comme tête, le critère sémantique de non-effacement. Les déterminants sont des prédicats qui prennent le nom comme argument : quand on dit *deux personnes*, cela signifie que 'personne' est au nombre de 'deux'. De même, *le chien* signifie que 'chien' est « défini », c'est-à-dire identifiable parmi les référents potentiels. En conséquence, puisque le déterminant est non effaçable, ce n'est pas un dépendant et il est donc une tête syntaxique.

9.28 Nom et déterminant comme co-têtes

Il est temps de conclure notre étude de la tête du groupe substantival. Comme on le voit, les données sont complexes et il est difficile de défendre que seul le nom ou seul le déterminant est la tête du groupe substantival. La troisième solution qui consiste à les considérer comme **co-têtes** est donc la plus séduisante, bien qu'elle oblige à manipuler une structure de dépendance qui n'est plus un arbre.

Notons encore que, si l'on veut vraiment un arbre de dépendance, le choix entre les deux arbres possibles — celui où le déterminant dépend du nom et celui ou le nom dépend du déterminant — n'est pas non plus fondamental. En effet, ce **choix** reste **local** dans l'arbre et que le passage d'une structure à l'autre est toujours possible et facile à réaliser, en inversant tout simplement les positions du nom et du déterminant. On peut alors voir ce choix comme une simple convention de notation : donner le statut de tête au déterminant privilégie les contraintes grammaticales, tandis que prendre le nom comme tête est un choix plus intuitif, car plus sémantique.

Bien que l'analyse comme co-tête nous semble la meilleure, dans la suite, il nous arrivera souvent de travailler avec un arbre de dépendance en choisissant souvent le nom comme tête du groupe substantival.

9.29 Co-occupation **Q**

Nous appelons co-occupation d'une position syntaxique la grammaticalisation d'un phénomène d'entassement. L'entassement, que nous étudierons en détail dans le Chapitre ??, est la possibilité pour un élément de venir s'entasser sur un autre élément et d'occuper ainsi la même position, comme dans une coordination.

Il nous semble que les propriétés particulières de la combinaison du nom et du déterminant, qui ont tous les deux des traits de tête, pourraient résulter d'un phénomène de co-occupation à l'origine. Rappelons que le latin n'a pas de déterminant et que les articles du français ont des origines diverses : *un* est un numéral et *le* vient du pronom démonstratif *ille* 'ça'. Le même pronom a d'ailleurs aussi

donné le pronom personnel *il/le/lui*. On peut donc postuler que la présence de l'article découle au départ d'une co-occupation de la position substantivale par un nom et un pronom : *le livre*, du latin *ille liber*, signifiant d'abord 'ça livre', chacun des deux éléments, *le* et *livre*, pouvant occuper seul la position en ancien français. La construction déterminant-nom serait donc clairement à l'origine une construction à deux têtes, de deux éléments venant occuper conjointement la même position. Puis l'usage aurait fini par imposer cette co-occupation jusqu'à rendre les deux éléments indissociables.

On peut se demander au passage jusqu'à quel point le pronom et l'article *le* sont devenus deux lexèmes distincts en français moderne. Considérons les données suivantes :

- (92) Je vois **le** livre rouge.
- (93) Je vois le rouge.
- (94) Je **le** vois.

Le pronom latin *ille* en se grammaticalisant a perdu de sa substance phonique et doit se cliticiser sur un l'élément à sa droite dans le groupe substantival (par *se cliticiser sur*, nous voulons dire 'former un groupe accentuel avec'; voir Chapitre ??). Si celui-ci est absent, il se place devant le verbe et peut ainsi continuer à se cliticiser sur l'élément qui le suit. Une telle explication amène à considérer que le pronom *le* du dernier exemple et le même objet que l'article *le* du premier exemple. Si tel est le cas, contrairement à ce qu'on avance habituellement, le deuxième exemple n'est pas elliptique, mais simplement un cas où le pronom gouverne un adjectif. (Lequel pronom ne pourrait pas d'ailleurs gouverner un complément de nom ou une relative, *le livre de syntaxe* donnant *celui de syntaxe* et non **le de syntaxe*, peut-être encore pour des problèmes de cliticisation, la préposition *de*, non accentuable, n'étant pas un hôte adéquat.)

On peut donner d'autres exemples de co-occupation potentielle. Le lexème *tout*, dont la syntaxe est si atypique, est un bon candidat :

(95) J'appelle tous mes amis.

Sa position devant le déterminant rend difficile d'en faire un adjectif. Il est parfois analysé comme un pré-déterminant, c'est-à-dire un élément qui se combinerait avec les déterminants définis (ainsi que *un* comme dans *Ils ont rasé tout un village*), mais cela n'expliquerait pas qu'il puisse également cooccurrer avec le pronom personnel comme dans :

(96) Je les appelle tous.

Il nous semble plus simple de considérer qu'il s'agit dans les deux cas d'un pronom qui vient co-occuper la position syntaxique d'un groupe substantival.

La conjonction de subordination *que*, comme la conjonction de subordination *that* pour l'anglais, semblent également relever d'une co-occupation au départ. La situation est un peu différente dans les deux langues. En français, la conjonction *que* est liée étymologiquement au pronom interrogatif *quoi* dont elle est la forme faible (c'est-à-dire non accentuée, d'où le changement de voyelle, le même qui fait passer de *moi* à *me*). On peut, en suivant Pierre Le Goffic, considérer que le *que* moderne reste davantage lié qu'on ne le pense habituellement à ce pronom interrogatif. En effet à côté de :

(97) Marie pense que Pierre dort.

on a:

(98) Marie pense quoi? Pierre dort.

On voit alors comment la complétive s'est construite : par co-occupation de la position d'objet du verbe penser par le pronom quoi, dans sa forme faible *que*, et par la proposition *Pierre dort*. Puis « *quoi Pierre dort* » s'est figé pour donner la forme moderne *que Pierre dort*, où les deux éléments, la conjonction *que* et le verbe fini, sont indissociables. En anglais, la conjonction est étymologiquement liée au démonstratif *that* 'ça', mais le phénomène est similaire :

- (99) Mary thinks that: Peter is sleeping. 'Mary pense ça: Peter dort.'
- (100) Mary thinks that Peter is sleeping. 'Mary pense que Peter dort.'

À la différence du français, la co-occupation n'est pas obligatoire et la proposition subordonnée peut occuper seule la position d'objet du verbe THINK 'penser': *Mary thinks Peter is sleeping*. Il n'y a pas lieu de parler ici d'effacement de la conjonction *that* (comme on le dit souvent), mais simplement de co-occupation optionnelle.

On peut penser que, dans la plupart des langues, certains marqueurs de position syntaxique sont, comme l'article pour le groupe substantival ou la conjonction pour la complétive, les versions grammaticalisées d'éléments qui à l'origine pouvaient occuper seuls la position.

9.30 Les grammèmes comme tête **Q**

Nous avons discuté du nom ou du déterminant comme tête du groupe substantival. Mais il existe une autre solution encore. Nous avons vu que la flexion du verbe pouvait être considérée comme la tête de la proposition et prendre le sujet comme dépendant (voir la section 9.14 sur *Le critère distributionnel sans effacement*). Une analyse similaire peut être proposée pour le nom. Les deux catégories flexionnelles du nom sont la définitude et le nombre (voir le Chapitre ?? sur les *Catégories flexionnelles*). On peut alors proposer pour *les deux amis* la structure suivante (dont les nœuds sont occupés par des syntaxèmes):

(101)

Si on ramène maintenant cette structure à une structure de dépendance entre mots, on a deux façons de la réduire :

- (102) soit on amalgame les grammèmes avec le lexème qui les marque :
 - soit on considère que ces grammèmes sont des flexions du nom et on les amalgame avec lui :

Comme on le voit, on retombe ainsi sur les deux arbres qu'on aurait obtenu en décidant soit que le déterminant est la tête, soit que le nom est la tête. Ceci est encore une façon de dire que le choix du nom ou du déterminant comme tête est finalement secondaire, puisqu'on peut voir ça comme deux réductions différentes d'une même analyse sous-jacente.

9.31 Dominance et projections maximales

La dépendance induit une autre relation : la DOMINANCE (voir l'encadré *Dépendance, dominance et transitivité* ci-dessous pour une définition formelle).

Nous dirons qu'un nœud B est dominé par le nœud A si B = A ou bien si B dépend de A ou bien si B dépend d'un dépendant de A et ainsi de suite. Autrement dit, B est dominé par A si B appartient au sous-arbre de dépendance dont A est la racine.

Prenons un exemple :

- 9 Tête et dépendance : Hiérarchiser la structure syntaxique
- (103) Beaucoup de gens aimeraient passer Noël en Laponie.

dont voici l'arbre de dépendance :

Figure 9.10 - Arbre de dépendance

Le sous-arbre de racine passer est :

de racine passer

Figure 9.11 - Sous-arbre

Les éléments dominés par le nœud *passer* sont donc *passer* lui-même, *Noël, en* et *Laponie*.

L'unité syntaxique formée par les nœuds dominés par un nœud A est appelée la PROJECTION MAXIMALE de A.

La projection maximale de *passer* est l'unité syntaxique *passer Noël en Laponie*. On dit simplement la PROJECTION de A lorsqu'on ne considère pas d'autres types de projections (voir la section 10.16 sur les *Projections partielles et ordre de saillance* pour une discussion sur les projections en général).

Dans la grammaire traditionnelle, les projections maximales sont appelées des GROUPES : GROUPE PRÉPOSITIONNEL (depuis ce matin), GROUPE ADJECTIVAL (très facile à comprendre), GROUPE ADVERBIAL (beaucoup trop vite), GROUPE SUBSTANTIVAL (la petite maison dans la prairie), etc. La projection maximale de la forme verbale est la PROPOSITION. Pour le terme de groupe verbal, dont nous ne faisons pas usage, nous renvoyons à l'encadré ?? éponyme.

Attention : dans beaucoup d'ouvrages de linguistiques, le terme *syntagme* est utilisé (en traduction du terme anglais *phrase*) pour désigner les groupes. Il est alors dit *syntagme nominal* (angl. *noun phrase*) pour *groupe nominal*. **Nous rejetons cet emploi du terme** *syntagme*, que nous utilisons dans son sens originel (celui donné par Saussure) d'unité obtenue par une com-



binaison syntagmatique.

Dans notre terminologie, groupe et syntagme sont deux notions bien distinctes : Marie forme un groupe substantival sans être pour autant un syntagme (il n'y a pas de combinaison), tandis que aimeraient passer est un syntagme, mais pas un groupe.

Si l'on adopte une représentation syntaxique très granulaire où les syntaxèmes flexionnels occupent des positions distinctes des lexèmes, on distinguera les PRO-JECTIONS FLEXIONNELLES des PROJECTIONS LEXICALES. Les projections flexionnelles sont généralement appelées les projections fonctionnelles dans la littérature en raison de la Syntaxe X-barre qui appelle catégories fonctionnelles les éléments qui ont uniquement une fonction grammaticale (voir l'encadré ?? sur la Syntaxe X-barre).

9.32 Dépendance, dominance et transitivité



En termes mathématiques, un arbre de dépendance représente une RELATION BINAIRE sur l'ensemble des mots, qu'on appelle la RELATION DE DÉPENDANCE. Intuitivement, une relation binaire est un appariement d'éléments deux à deux. Donnons une définition plus formelle.

Si E est un ensemble, on note $E \times E$ l'ensemble des couples sur E. Par exemple, si E = $\{a,b\}$, alors E × E = $\{(a,a), (a,b), (b,a), (b,b)\}$. On appelle relation binaire sur E tout sous-ensemble de E × E. Si R est une relation binaire sur E, on privilégie la notation a R b pour dire que les éléments a et b de E sont en relation par R (plutôt que de dire que le couple (a,b) est élément de R). Toute relation binaire peut être représentée par un graphe orienté dont les nœuds sont les éléments de E et une arête de a vers b représente la relation entre les nœuds a et b (a R b). La figure suivante représente le Graphe de la relation $R = \{ (1,2), (2,1), (1,3), (3,3) \}$

(104)

L'arbre de dépendance est ainsi le graphe de la relation de dépendance. Dans l'exemple de la section précédente, $R = \{(gens, de), (de, beaucoup), (beaucoup, aimeraient), \}$ (passer,aimeraient), (Noël,passer), (en, passer), (Laponie,en) \}.

Voyons maintenant comment définir la RELATION DE DOMINANCE. Avant cela, nous devons présenter quelques propriétés typiques des relations que nous exploiterons.

Une relation binaire R sur E est dite réflexive si tout élément de E est en relation avec lui-même (pour tout a dans E, on a a R a). Une relation binaire R sur E est dite symétrique si à chaque fois que a R b, on a aussi b R a. Une relation binaire R sur E est dite transitive si à chaque fois que a R b et b R c, on a aussi a R c. La réflexivité, la symétrie et la transitivité correspondent aux configurations suivantes :

(105) Réflexivité Symétrie Transitivité

La relation de dépendance n'est ni réflexive, ni symétrique, ni transitive. Elle est même à l'opposé de ça, puisqu'on n'a jamais une configuration des types précédents dans un arbre de dépendance. Mais on peut construire à partir de la relation de dépendance une relation réflexive et transitive, et c'est justement la relation de dominance.

La relation de dépendance. En d'autres termes, il s'agit de la plus petite relation transitive et réflexive qui contienne la relation de dépendance. Elle est obtenue en ajoutant les configurations correspondant à la transitivité et la réflexivité sur l'arbre de dépendance jusqu'à saturation (c'est en ce sens qu'il s'agit d'une *clôture* pour la transitivité et la réflexivité).

Formellement, la clôture transitive et réflexive d'une relation binaire R est la relation R* définie par : a R* b s'il existe n+1 nœuds a_0 , a_1 ,..., a_n (pour un entier quelconque $n \ge 0$) tels que $a = a_0$, $b = a_n$ et a_i R a_{i+1} pour tout i compris entre 0 et n-1. Autrement dit, a R* b s'il existe un chemin orienté de a à b dans le graphe de la relation R (de longueur éventuellement nulle pour n = 0).

De même que nous avons construit la relation de dominance à partir de la relation de dépendance, on peut construire la relation de dépendance à partir de la relation de dominance. La relation de dominance est non seulement réflexive et transitive, mais elle est aussi antisymétrique, comme l'est la relation de dépendance. L'antisymétrie est un cas particulier de l'acyclicité. Une relation est antisymétrique si la configuration correspondant à la symétrie n'apparaît jamais, c'est-à-dire s'il n'existe pas de nœuds a et b distincts avec a R b et b R a.

Une relation réflexive, antisymétrique et transitive est appelée une RELATION D'ORDRE. Il s'agit d'un ordre au sens où on l'entend pour les nombres quand on dit « $3 \le 7$ » : la relation \le est une relation d'ordre sur les nombres entiers. Tout

relation d'ordre ordonne les éléments en déclarant certains éléments comme plus grand que d'autres.

La relation de dominance est une relation d'ordre. Rappelons qu'elle étend la relation de dépendance dont le graphe est l'arbre de dépendance. La racine de cet arbre domine tous les autres nœuds et est donc le plus grand élément, tandis que les feuilles sont les plus petits éléments. La relation de dominance est une relation d'ordre partielle : elle est partielle dans le sens où tous les nœuds ne sont pas ordonnés les uns par rapport aux autres. Par exemple, si deux nœuds sont frères dans l'arbre de dépendance (c'est-à-dire qu'ils ont le même gouverneur), alors aucun des deux ne domine l'autre.

Toute relation d'ordre sur un ensemble fini induit une RELATION DE PRÉCÉDENCE. Le prédécesseur de 5 pour la relation ≤ est 4, c'est-à-dire le plus grand des nombres strictement plus petit que 5. Plus généralement, un prédécesseur de l'élément *a* pour une relation d'ordre quelconque est un plus grand élément parmi les éléments plus petits que *a*. Pour un ordre partiel, un élément peut avoir plusieurs prédécesseurs (et plusieurs successeurs). Pour la relation de dominance, le successeur d'un nœud est son gouverneur et ses prédécesseurs sont ses dépendants. La relation de dépendance est donc la relation de précédence induite par la relation de dominance.

En conclusion, la donnée d'une relation de dépendance ou d'une relation de dominance revient au même. Il s'agit de deux façons équivalentes de définir une même hiérarchie. Chacune des deux relations est immédiatement déductible de l'autre.

9.33 Critère de recouvrement

Une dernière façon de caractériser une tête repose sur l'organisation linéaire ou topologie (traditionnellement appelée l'ordre des mots) (voir Chapitre 11). Le test de recouvrement est un cas particulier de la propriété de projectivité que nous présenterons à la section 11.15 *Projectivité et dépendance projective*. L'idée générale est que les dépendants tendent à se placer à côté de leur gouverneur. Cela n'est pas toujours possible si le gouverneur B a beaucoup de dépendants et il est donc possible qu'un autre dépendant de B se place entre B et son dépendant C. Mais par contre, il n'est généralement pas possible que le gouverneur de B se place entre B et son dépendant C.

CRITÈRE DE RECOUVREMENT. Si A est **entre** B et C et que B et C sont connectés, alors il est peu probable que A gouverne B. Ou dit autrement, si A gouverne B, alors la connexions entre B et un de ses dépendants C ne peut généralement **pas recouvrir** A.

Figure 9.12 - Configuration rejetée par le critère de recouvrement

Illustrons cela avec la relation sujet-verbe et considérons les données suivantes :

- (106) Le chien blanc dort.
- (107) *Le chien dort blanc.
- (108) Le chien dort depuis ce matin.
- (109) Depuis ce matin, le chien dort.

Nous testons ici la connexion entre A = dort et B = le chien. On voit que la dépendance entre B et son dépendant C = blanc ne peut pas recouvrir A. Ceci est un indice du fait que dort gouverne le chien obligeant le chien blanc à être entièrement d'un des deux côtés du verbe. À l'inverse, le fait de pouvoir placer le complément depuis ce matin qui dépend de dort à gauche de le chien est un indice du fait que le chien ne gouverne pas dort (ou alors nous aurions affaire à un cas de non-projectivité).

Ce test ne fonctionne qu'avec certaines constructions. Même dans le cas que nous venons d'étudier, il est possible de violer la projectivité et de placer un dépendant du sujet au delà du verbe (voir l'exemple *une belle est entrée qui voulait les acheter* que nous étudions dans la section 11.15).

9.34 Multiplicité des analyses en dépendance

Arrivé à la fin de ce chapitre qui présente l'outil principal de notre analyse syntaxique, à savoir la dépendance, le lecteur aura noté que nous ne définissons pas une unique structure de dépendance pour chaque énoncé, mais un ensemble de structures possibles. Nous avions déjà évoqué la nécessité de pouvoir considérer

une analyse à différents niveaux de granularité (voir la section 8.18 sur *Structures de connexion, granularité et critères*). Une dépendance ne lie pas seulement deux mots ou deux syntaxèmes : elle lie deux côtés de l'énoncé et ces deux côtés peuvent être appréhendés avec plus ou moins de précision. Nous continuerons à développer ce point au chapitre suivant lorsque nous introduirons les constituants.

Le passage de la connexion à la dépendance, en raison de l'abondance de critères définissant la tête, introduit de nouveaux paramètres de variation de la structure : le fait de favoriser tel ou tel critère peut amener à des choix de représentations syntaxiques légèrement différents (voir l'encadré ?? Quelle structure pour quels critères?). Ces choix dépendent généralement des objectifs : pédagogique (proposer une représentation graphique permettant de communiquer avec l'élève, l'étudiant ou le collègue), pratique (avoir une analyse qui puisse servir de base à des calculs permettant par exemple de simuler l'analyse ou la synthèse d'énoncés) ou cognitif (tenter de modéliser ce qui se passe dans le cerveau d'un locuteur lorsqu'il parle).

L'objectif cognitif est séduisant, mais il est actuellement inaccessible : nous n'avons pas les moyens de prouver que telle ou telle analyse est bien celle que fait notre cerveau, ni même de prouver que notre cerveau construit des structures, et encore moins que notre cerveau construit des structures comme celles que nous présentons. Nous pouvons néanmoins, à l'intérieur d'un modèle formel d'une langue donnée, montrer que le choix d'une structure plutôt qu'une autre permet de faire de meilleures prédictions sur les capacités de production des locuteurs.

L'objectif pédagogique nous pousse, lui, à pouvoir construire différentes représentations. Il est souvent nécessaire de ne pas présenter une analyse dans toute sa complexité et d'utiliser des représentations facilement interprétables. C'est certainement ce qui amène de nombreux auteurs à n'utiliser que des arbres pour leurs analyses, même s'il y a ici et là des indices que les connexions ne sont pas toujours si clairement hiérarchisées.

L'objectif pratique amène aussi à rechercher la simplicité. Inutile d'introduire plus de détails que nécessaire pour les calculs qu'on envisage. Ici encore, de nombreux chercheurs en traitement automatique vont privilégier des arbres, structures pour lesquelles les algorithmes sont bien mieux maîtrisés.

Si l'on peut donc associer différentes structures à un même énoncé, il faut aussi comprendre que, à l'inverse, **une même structure peut avoir différentes inter- prétations**. La « sémantique » associée à une dépendance n'est en rien univoque : elle dépend complètement des critères qui servent à définir la tête. Quand nous postulons une dépendance entre *une foule* et *de gens* dans *une foule de gens ap-*

plaudissent (voir l'encadré ?? sur les Déterminants complexes) que voulons-nous dire ? Que une foule contrôle la valence supérieure de une foule de gens ? Que une foule régit de gens ? Il n'est même pas toujours sûr qu'une dépendance donnée vérifie bien les critères qu'on a retenus : elle peut résulter de compromis généraux, par exemple sur la nécessité d'avoir une structure arborescente pour nos calculs.

Cela ne doit pas donner l'impression que les structures que nous proposons sont arbitraires. Il existe clairement de mauvaises analyses des phénomènes syntaxiques: par exemple, des analyses qui oublieraient des connexions qui répondent clairement aux critères que nous nous sommes fixés ou qui en ajouteraient qui n'y répondent pas. Par ailleurs, même s'il existe des cas problématiques dont l'analyse peut varier en fonction des critères retenus, la plupart des connexions sont indiscutables et clairement hiérarchisées. Et lorsque plusieurs solutions semblent possibles, on peut souvent montrer que certaines structures sont plus simples, moins redondantes que d'autres tout en ayant le même contenu informationnel et donc faire des choix bien argumentés.

Dans les chapitres suivants, il nous arrivera de faire varier la représentation choisie en fonction de la nature des phénomènes étudiés et de nos exigences pédagogiques.

9.35 Dépendance, constituance et topologie

Nous avons maintenant terminé la définition de ce qu'est une structure de dépendance en introduisant d'abord la connexion (Chapitre 8), puis en montrant comment hiérarchiser les connexions en dépendances en introduisant la notion de tête (ce chapitre). Dans le prochain chapitre (3.4), nous montrerons qu'un arbre de dépendance induit naturellement un arbre de constituants et nous étudierons l'équivalence entre les deux types de structure. Ceci nous amènera à présenter d'autres méthodes pour définir la structure de dépendance. À la fin du Chapitre 10, nous verrons comment construire la structure syntaxique d'exemples complexes en combinant différentes méthodes. Certaines constructions, comme la coordination ou la relativisation, feront l'objet de chapitres séparés dans la partie 5 sur la *Microsyntaxe*, qui est le domaine de la rection et donc le domaine d'excellence de la dépendance. Le Chapitre 11 sera consacré à la linéarisation de la structure syntaxique et à la structure topologique qui en découle. Le Chapitre ?? présentera les liens entre structure syntaxique et structure sémantique au travers de la structure syntaxique profonde.



Exercices

Exercice 1 Nous donnons ci-après une analyse faite par Claude Buffier au début du 18^e siècle. Nous vous proposons de traduire graphiquement cette analyse à l'aide d'un diagramme. Vous représenterez les relations entre les unités syntaxiques par des flèches.

« Un homme qui étourdit les gens qu'il rencontre avec de frivoles discours a coutume de causer beaucoup d'ennui à tout le monde. Je dis que dans ce discours, tous les mots sont pour modifier le nom un homme, et le verbe a coutume, et que c'est en cela que consiste tout le mystère et toute l'essence de la syntaxe des langues : 1° le nom un homme, est modifié d'abord par le qui déterminatif : car il ne s'agit pas ici d'un homme en général, mais d'un homme marqué et déterminé en particulier par l'action qu'il fait d'étourdir; de même il ne s'agit pas d'un homme qui étourdit en général, mais qui étourdit en particulier les gens, et non pas les gens en général, mais en particulier les gens qu'il rencontre. Or cet homme qui étourdit ceux qu'il rencontre, est encore particularisé par avec des discours, et discours est encore particularisé par *frivoles*. On peut voir le même dans la suite de la phrase : a coutume est particularisé par de causer, de causer est particularisé par ses deux régimes, par son régime absolu, savoir, beaucoup d'ennui, et par son régime respectif, à tout le monde. Voilà donc comment tous les mots d'une phrase quelque longue qu'elle soit, ne sont que pour modifier le nom et le verbe. » (Buffier1709, p. 84, nous modernisons l'orthographe)

Exercice 2 Nous donnons ci-après une analyse faite par Claude Chesneau dit Du Marsais dans l'article « Construction » de l'*Encyclopédie* de Diderot et d'Alembert. Nous vous proposons de traduire graphiquement cette analyse à l'aide d'un diagramme. Vous représenterez les unités syntaxiques considérées par des bulles ; les termes relationnels seront associés à des flèches entre les bulles et les termes non relationnels seront traités comme des étiquettes des bulles.

- « Celui qui me suit, dit Jésus-Christ, ne marche point dans les ténèbres : considérons d'abord cette phrase ou cet assemblage de mots grammatica-lement, c'est-à-dire selon les rapports que les mots ont entre eux; rapports d'où résulte le sens : je trouve que cette phrase, au lieu d'une seule proposition, en contient trois.
- 1°. Celui est le sujet de ne marche point dans les ténèbres; et voilà une proposition principale; celui, étant le sujet, est ce que les Grammairiens appellent le nominatif du verbe. Ne marche point dans les ténèbres, c'est l'attribut; [...] ne point est la négation, qui nie du sujet l'action de marcher dans les ténèbres.

Dans les ténèbres, est une modification de l'action de celui qui marche, il marche dans les ténèbres; dans est une préposition qui ne marque d'abord qu'une modification ou manière incomplète; c'est-à-dire que dans étant une préposition, n'indique d'abord qu'une espèce, une sorte de modification, qui doit être ensuite singularisée, appliquée, déterminée par un autre mot, qu'on appelle par cette raison le complément de la préposition : ainsi les ténèbres est le complément de dans; et alors ces mots, dans les ténèbres, forment un sens particulier qui modifie marche, c'est-à-dire qui énonce une manière particulière de marcher.

2°. *Qui me suit*, ces trois mots font une proposition incidente qui détermine *celui*, et le restreint à ne signifier que le disciple de Jésus-Christ, c'est-à-dire celui qui règle sa conduite et ses mœurs sur les maximes de l'Evangile : ces propositions incidentes énoncées par *qui*, sont équivalentes à un adjectif.

Qui est le sujet de cette proposition incidente ; *me suit* est l'attribut ; *suit* est le verbe ; *me* est le déterminant ou terme de l'action de *suit* : [...]

3°. Dit Jésus-Christ, c'est une troisième proposition qui fait une incise ou sens détaché; c'est un adjoint : en ces occasions la construction usuelle met le sujet de la proposition après le verbe : Jésus-Christ est le sujet et dit est l'attribut. » (Dumarsais, article « Construction », Encyclopédie, vol. 4, 1754)

Exercice 3 Analyse distributionnelle. Montrer que les formes verbales infinitives ont une valence supérieure proche de celle des groupes substantifs. Montrer qu'il existe néanmoins des différences (on pourra regarder les

contextes proposés par des verbes comme vouloir, pouvoir ou tenter).

Exercice 4 Dans son article de 1977 (p. 8), Paul Garde propose les trois critères suivants pour considérer que B est la tête de l'unité AB :

- « 1) dans un contexte donné, A peut être supprimé, non B :
- $X (AB) \rightarrow X (B)$ et non *X (A)
- 2) dans un contexte donné, A peut être remplacé par autre chose, B ne le peut pas :
- $X (AB) \rightarrow X (A'B)$ et non *X (AB')
- 3) une modification du contexte entraı̂ne un changement de B, non de A :
- $X (AB) \rightarrow Y (AB')$ et non *Y (A'B) »

Montrer qu'il s'agit, dans l'intention, des critères distributionnels avec et sans effacement, mais que, pour être corrects, les critères doivent être énoncés et formalisés de manière légèrement différente.

Exercice 5 On considère les données suivantes :

- (110) Je prends les deux livres.
- (111) Je prends les deux.
- (112) Je les prends tous les deux.
- (113) Je les prends tous.
- (114) ?* Je les prends les deux. (possible uniquement avec un détachement de les deux)
- (115) *Je prends tous les deux.

Ces données nous permettent-elles de statuer sur la tête de l'unité *tous les deux*?

Exercice 6 Montrer que le nom Fois possède une valence supérieure différente des autres noms.

Exercice 7 a) Montrer que dans une forme verbale composée comme *a dormi*, c'est le verbe auxiliaire qui régit le participe passé et non l'inverse.

b) Pourquoi ne retenons-nous pas le fait que certains verbes, comme ARRIVER, prenne l'auxiliaire ÊTRE et non AVOIR au passé composé comme un critère pertinent pour décider qui de l'auxiliaire ou du participe est la tête de la forme verbale composée?

Exercice 8 La distribution des formes verbales infinitives du français est différente de celle des formes finies. Ceci plaide pour considérer la flexion comme la tête externe des formes verbales. Le syntaxème d'infinitif est-il aussi la tête interne de la forme verbale infinitive?

Exercice 9 On étudie les groupes substantifs de la phrase suivante :

(116) Le journal Libération a publié un article sur l'affaire Dreyfus.

Comment déterminer la tête des groupes *le journal Libération* et *l'affaire Dreyfus*? Montrer que seuls les critères topologiques et sémantiques s'appliquent réellement.

Exercice 10 On s'intéresse à la construction $de\ X\ a\ Y$, illustrée par des exemples comme $le\ train\ de\ Paris\ a\ Marseille$ ou $il\ travaille\ du\ lundi\ au\ vendredi$. À quelle hypothèse nous conduit l'application du test de déplacement? Cette hypothèse est-elle confirmée par d'autres tests? Comment analyser alors la construction?

Exercice 11 Construire la structure de dépendance de n'importe quel énoncé dans n'importe quelle langue constitue un exercice d'application de ce chapitre. Nous proposons de chercher la structure syntaxique des exemples suivants :

(117) a. J'y suis particulièrement sensible.

- b. Un tiers des participants seront éliminés dès que le jeu commencera.
- c. Ma voiture n'a pas voulu démarrer ce matin.
- d. Le professeur est arrivé deux minutes après la sonnerie.
- e. Ma chemise a l'air moins blanche qu'avant.
- f. Epuisée par sa course, Zoé a préféré s'arrêter pour souffler un peu.
- g. Seuls mes amis peuvent accepter une telle demande.

Chacun de ces exemples un problème de syntaxe que l'on tentera d'identifier et de résoudre. D'autres problèmes plus complexes, comme la coordination ou les pronoms relatives, seront présentés dans les chapitres 5.3 et 5.4.

Exercice 12 La construction causative en faire Vinf (Vinf = verbe à l'infinitif) du français est particulièrement problématique. Elle possède deux propriétés remarquables.

Premièrement, dans cette construction, le sujet du verbe à l'infinitif introduit par faire est rétrogradé et sa fonction dépend de la valence du verbe. Le sujet retrogradé peut être :

- soit complément d'objet direct : Zoé fait dormir les enfants.
- soit complément d'objet indirect : Zoé fait manger la soupe aux enfants.
- soit complément oblique (dit d'agent) : Zoé fait porter un livre à Luc par les enfants.

Deuxièmement, les compléments de la construction se cliticisent devant le verbe faire qu'il s'agisse du sujet retrogradé ou des compléments régis par le Vinf : *Zoé la leur fait manger*.

La question est de savoir si le sujet retrogradé dépend de faire ou du Vinf. Voyez quels sont les critères qui s'appliquent et conclure.

Exercice 13 En allemand, les prépositions locatives peuvent être suivies d'un groupe substantival au datif ou à l'accusatif. Le datif donne une interprétation statique au groupe prépositionnel, tandis que l'accusatif lui

donne une interprétation dynamique/directionnelle. Ceci est illustré par la paire suivante :

(118) a. Helga joggt oft in **diesem** Stadion.

Helga fait souvent du footing dans ce.DAT stade.

fait ... footing is discontinous. Maybe choose other verb?

b. Helga joggt oft in dieses Stadion.
 Helga fait souvent du footing (qui se termine) dans
 ce.ACC stade.

Certains verbes néanmoins ne permettent qu'un des deux cas :

- (119) a. Wolfgang ist in der Schule. Wolfgang est à la.dat école.
 - b. *Wolfgang ist in die Schule. Wolfgang est à la.Acc école.
 - c. Thomas zählt auf seine Mutter. Thomas compte sur sa.Acc mère.
 - d. #?? Thomas zählt auf seiner Mutter.

 Thomas compte sur sa.DAT mère
 signifierait au mieux 'Thomas compte en étant assis sur sa mère.'

Qu'est que cela nous dit sur la tête du groupe prépositionnel?

Exercice 14 Construction génitive. En allemand, le complément de nom est au génitif (au lieu d'être marqué par une préposition *de* comme en français) : *das Haus Olafs* 'la maison d'Olaf', lit. la maison Olaf.GEN. Mais ce complément génitif peut aussi venir à la place du déterminant comme en anglais : *Olafs Haus* vs. *Olaf's house* 'la maison d'Olaf'. Quels arguments ces constructions donnent-elles pour ou contre le déterminant comme tête ? Comment les analyseriez-vous ?

9.35 Dépendance, constituance et topologie



Lectures additionnelles

Pour une présentation très complète des critères définissant la tête d'une dépendance syntaxique, on consultera Mel'čuk1988 ou le volume 2 de Mel'čuk & Milićević (??), ainsi que l'article de Garde1977 (voir exercice 4). Sur la notion de tête, Zwicky1985 reste une référence incontournable, que l'on doit compléter par la réponse donnée par Hudson1987. Les ouvrages de Richard Hudson sur sa Word Grammar et notamment Hudson2007 constituent de bonnes introductions à la dépendance.

Il existe de nombreux écrits sur le choix entre nom ou déterminant comme tête du groupe substantival, parmi lesquels **Hudson1984**, **Abney1987**, Van **Langendonck1994** et **Hudson2007**.

Sur la rection faible, rapidement évoquée en 3.3.19, on pourra consulter Blanche-Benveniste1989. Le fait que certaines prépositions sont des têtes faibles est défendu dans la thèse de Tseng2001.

Pour remonter aux origines de l'analyse en dépendance, on consultera **Tesnière1959** et **Tesnière1934**. Pour la notion de tête, voir **Bloomfield1933**. Pour la notion de valence, l'article de **Peirce1897** et **Tesnière1959**. Pour la notion de valence supérieure, voir **Mel'čuk1988**.

Ceux qui s'intéressent aux écrits anciens pourront trouver les ouvrages de Buffier, Girard ou Beauzée sur le site de la Bibliothèque Nationale de France (gallica.bnf.fr) ou sur Google Books. L'*Encyclopédie* de Diderot et D'Alembert, avec les articles de Dumarsais et Beauzée, est disponible sur enccre.academie-sciences.fr. **Kouloughli1999** discute de la notion de dépendance dans les grammaires arabes et montre qu'il s'agit davantage de rection morphologique que syntaxique, ce qui rejoint le point de vue de **Chevalier1968** qui date la version contemporaine de la notion de complément des travaux de Beauzée. Il existe diverses ressources en ligne concernant les diagrammes de Reed & Kellogg et la grammaire de Reed & Kellogg est régulièrement rééditée!

Les actes de langage et la fonction illocutoire sont introduits dans les ouvrages de Gardiner1932, Bloomfield1933, Austin1962 et Searle1969.

Steven Abney1987 The English Noun Phrase in Its Sentencial Aspect, PhD

thesis, Cambridge: MIT.

John L. Austin1962 *How to Do Things with Words*, Cambridge : Harvard University Press, [traduction française, 1991, *Quand dire, c'est faire*, Seuil].

Nicolas Beauzée1767 Grammaire générale ou Exposition raisonnée des éléments nécessaires du langage, pour servir de fondement à l'étude de toutes les langues, vol. 2 : Syntax, Paris : Barbou.

Claire Blanche-Benveniste1989 Constructions verbales "en incise" et rection faible des verbes, *Recherches sur le français parlé*, 9, 53-74.

Leonard **Bloomfield1933** *Language*, New York, [traduction française, 1969, *Le langage*, Paris : Payot].

Claude **Buffier1709** *Grammaire françoise sur un plan nouveau*, Paris : Le Clerc-Brunet-Leconte & Montalant.

Jean-Claude **Chevalier1968** *Histoire de la syntaxe : Naissance de la notion de complément dans la grammaire française (1530-1750)*, Droz.

Denis Diderot, Jean Le Rond D'Alembert (1751-1772) *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, 28 volumes. Edition en ligne : enccre.academie-sciences.fr.

Paul **Garde1977** Ordre linéaire et dépendance syntaxique : contribution à une typologie, *Bulletin de la Sociéte de Linguistique de Paris*, 72(1),1-26.

Alan **Gardiner1932** *Speech and Language*, Londres : Oxford.

Gabriel **Girard1747** Les Vrais principes de la langue françoise, ou la Parole réduite en méthode, Paris : Le Breton.

Richard Hudson1984 Word Grammar, Oxford : Blackwell.

Richard **Hudson1987** Zwicky on heads. *Journal of linguistics*, 23(1), 109-132.

Richard **Hudson2007** *Language Networks - The New Word Grammar*, Oxford University Press.

Otto **Jespersen1924** *The Philosophy of Language*, Londres : Allen & Unwin.

Otto Jespersen1937 Analytic Syntax, Londres: Allen & Unwin.

Franz ${\bf Kern 1883} \ Zur \ Methodik \ des \ deutschen \ Unterrichts,$ Berlin : Nicolai.

Djamel Eddine **Kouloughli1999** Y a-t-il une syntaxe dans la tradition arabe?, *Histoire Épistémologie Langage* 21(2), 45-64.

Igor Mel'čuk (??) Dependency Syntax - Theory and Practice, Albany : SUNY.

Igor Mel'čuk, Jasmina **Milićević2014** *Introduction à la linguistique*, volume 2, Paris : Hermann.

Charles S. Peirce1897 The logic of relatives, *The Monist*, 7, 161–217.

John **Searle1969** *Speech Acts*, Cambridge University Press, [traduction française, 1969, *Les actes de langage*, Hermann.

Roch-Amboise **Sicard1801** Élémens de grammaire générale appliqués à la langue française (2 volumes), Paris : Deterville.

Henry $\mathbf{Sweet1891}$ A new English Grammar, logical and historical, Oxford : Clarendon.

Lucien **Tesnière1934** Comment construire une syntaxe, *Bulletin de la Faculté des Lettres de Strasbourg*, 7, 12^{ème} année, 219-229.

Lucien **Tesnière1959** Éléments de syntaxe structurale, Paris : Klincksieck. Jesse **Tseng2001** *The Selection and Representation of Prepositions*, PhD thesis, University of Edinburgh.

Willy Van Langendonck1994 Determiners as heads?, *Cognitive Linguistics*, 5(3), 243-259.

Henri **Weil1844** *De l'ordre des mots dans les langues anciennes comparées aux langues modernes*, Paris : Crapelet.

Arnold Zwicky1985 Heads, Journal of Linguistics 21(1), 1-29.



Citations originales

Citations de la section ?? Jespersen (Jespersen1924 : 96) :

In any composite denomination of a thing or person [...], we always find that there is one word of supreme importance to which the others are joined as subordinates. This chief word is defined (qualified, mo-

dified) by another word, which in its turn may be defined (qualified, modified) by a third word, etc. We are thus led to establish different "ranks" of words according to their mutual relations as defined or defining. In the combination extremely hot weather the last word weather, which is evidently the chief idea, may be called primary; hot, which defines weather, secondary, and extremely, which defines hot, tertiary.

Sweet (Sweet1891: 16): Adjunct-words and Head-words.

40. The most general relation between words in sentences from a logical point of view is that of adjunct-word and head-word, or, as we may also express it, of modifier and modified. Thus in the sentences tall men are not always strong, all men are not strong, tall, strong, and all are adjunct-words modifying the meaning of the head-word men. So also dark, quick, quickly are adjunct-words in dark red, he has a quick step, he walks quickly. Stone is an adjunct-word in stone wall, wall of stone, because it modifies (defines) the meaning of wall. So also book (books) is an adjunct-word in book-seller, bookselling, sale of books, he sells books, he sold his books, the corresponding head-words being seller, selling, sale, sells, sold.

41. The distinction between adjunct-word and head-word is only a relative one: the same word may be a head-word in one sentence or context, and an adjunct-word in another, and the same word may even be a head-word and an adjunct-word at the same time. Thus in *he is very strong*, *strong* is an adjunct-word to *he*, and at the same time head-word to the adjunct-word *very*, which, again, may itself be a head-word, as in *he is not very strong*.

Bloomfield (Bloomfield1933: section 12.10):

Every syntactic construction shows us two (or sometimes more) free forms combined in a phrase, which we may call the *resultant* phrase. The resultant phrase may belong to a form-class other that of any constituent. For instance, *John ran* is neither a nominative expression (like *John*) nor a finite verb expression (like *ran*). Therefore we say that the English actor-action construction is *exocentric*: the resultant phrase belongs to the form-class of no immediate constituent.

On the other hand, the resultant phrase may belong to the same formclass as one (or more) of the constituents. For instance, *poor John* is proper-noun expression, and so is the constituent *John*; the forms *John* and *poor John* have, on the whole, the same functions. Accordingly, we say that the English character-substance construction (as in *poor John*, *fresh milk*, and the like) is an *endocentric* construction.

check the use of colons and quotations marks in all multiline quotations in book

Citations de la section ?? Kern (1883 : 10) :

Die Abhängigkeit eines Wortes von einem andern bezeichnet man (graphisch, durch ein Schema) durch einen von dem regierenden Wort nach unten gezogenen Strich, an dessen Ende das regierte (oder abhängige) Wort steht [figure] oder ohne Worte durch bloße grammatische Bezeichnungen [figure].



Corrections des exercices

Corrigé 1 Nous proposons de traduire l'analyse de Buffier par le diagramme suivant où les flèches représentent les relations du type « est déterminé par » ou est « est particularisé par ». Notons que, à l'exception du cas de *avec des discours*, les deux termes de la relations sont à chaque fois clairement donnés par Buffier. Nous représentons pour cette raison la

relation entre avec des discours et son gouverneur en pointillé.

(120) Diagramme d'après Buffier1709

Corrigé 2 Du Marsais propose une analyse hybride entre dépendance et constituance immédiate. Il semble hésiter entre les deux approches et utilise notamment le même terme *sujet* pour désigner la relation entre le groupe sujet et le groupe verbal (*Celui* est le sujet de *ne marche point dans les ténèbres*) et pour désigner un constituant de la proposition (*Qui me suit*, ces trois mots font une proposition incidente [...] *Qui* est le sujet de cette proposition incidente).

(121) Diagramme d'après Dumarsais1754

Corrigé 3 Le fait que les infinitifs commutent avec des groupes substantifs est déjà souligné par Buffier : « Ce que les Grammaires appellent communément des verbes, comme aimer, lire, dormir, font de véritables noms substantifs ; bien qu'ils aient des propriétés particulières. Car enfin ils signifient un sujet dont on peut parler ; ils sont souvent le nominatif [= sujet] des verbes et même leur régime [= complément] ; ils sont par conséquent de vrais noms. En effet quand on dit, Avouer sa faute est la réparer ; avouer est ici le sujet dont on parle, et se trouve le nominatif du verbe est. Et quand on dit je veux avouer ma faute : avouer est le véritable régime du verbe je veux ; comme ma faute est le régime de avouer. » (Buffier1709, p. 59)

Néanmoins, la commutation n'est pas toujours possible. Il existe des verbes comme pouvoir qui peuvent très difficilement prendre un substantif comme complément (*Elle peut voyager* vs *Elle peut tout* vs [?] *Elle peut ça* vs **Elle peut un voyage*). Par ailleurs, si avec certains verbes, les substantifs commutent avec un infinitif nu (*Elle veut un voyage* vs *Elle veut voyager* vs **Elle veut de voyager*), d'autres imposent un infinitif précédé de la préposition de (*Elle tente un voyage* vs **Elle tente voyager* vs *Elle tente de voyager*). Dans la position sujet, les deux formes sont possibles (*Voyager est un plaisir* vs *De voyager est plaisir*).

Corrigé 4 Le critère 1 de Paul Garde vise le critère distributionnel avec effacement. Il faut néanmoins préciser a) que le critère doit être valable pour

tout contexte (l'ensemble des X constitue alors la distribution de l'unité considérée); b) que A doit être supprimable dans au moins un contexte; c) que la réciproque doit être vraie. Par ailleurs, il n'est pas nécessaire de spécifier que B ne peut être supprimé (voir le cas des têtes effaçables dans l'encadré ??). En adoptant les notations de Garde, on peut donc reformuler le critère de la façon suivante :

(122) 1a) si A est supprimable dans un contexte, il doit l'être dans tous les contextes :

pour tout X, X (AB) \rightarrow X (B)

1b) inversement, A doit pouvoir être ajouté à B dans tous les contextes

pour tout X, X (B)
$$\rightarrow$$
 X (AB).

Les critères 2 et 3 visent le critère distributionnel sans effacement. La quantification sur les contextes doit être précisée comme précédemment. Le critère 2 dit deux choses : d'une part, que A peut être remplacé par un A' dans n'importe quel contexte (à condition évidemment que A' puisse se combiner avec B dans au moins un contexte) :

(123) pour tout A', s'il existe Y tel que Y (A'B), alors, pour tout X, X (AB) \rightarrow X (A'B).

Et d'autre part, que cette propriété n'est pas vraie pour B, c'est-à-dire qu'il doit exister au moins un contexte où B ne peut pas être remplacé par un B' (qui peut pourtant se combiner avec A) :

(124) il existe B', Y et X tels que Y (AB'), X (AB) et *X (AB').

Le critère 3 dit qu'(au moins) une modification du contexte entraîne un changement de B, ce qui devrait être formalisé par :

(125) il existe Y et B' tel que *Y (AB) et Y (AB')].

Dire que la même propriété n'est pas vraie pour A se formalise par :

(126) pour tout Y et A', *Y (AB) \rightarrow *Y(A'B).

Le critères 2 et 3 visent bien à dire que B contrôle la distribution de AB, c'est-à-dire qu'il existe un B' tel que AB et AB' aient des distributions différentes. Il faut néanmoins bien préciser que B' commute avec B et que donc B et B' s'excluent mutuellement :

(127) pour tout X, *X (ABB').

Même chose avec A et A'.

Corrigé 5 On voit que dans le contexte Je les prends_, U = tous les deux peut commuter avec A = tous, mais pas B = les deux. Par ailleurs, dans le contexte Je prends_, B est possible, mais pas U ou A. Autrement dit, U et A ont la même distribution, différente de B. Toutes ces données concordent pour désigner A = tous comme la tête de U.

Corrigé 6 Le nom fois combiné avec un déterminant se comporte comme un groupe adverbial (*Il vient en avril et en mai* vs *Il vient plusieurs fois*) et plus difficilement comme un groupe substantif, mais s'il peut être sujet (*Une fois suffit*; *Cette fois sera la bonne*). Voir la section 9.25 *Nom comme tête*?.

Corrigé 7 a) Nous avons que dans la section 9.14 sur *Le critère distributionnel sans effacement* que la flexion verbale contrôle la distribution de la forme verbale. Pour la même raison, l'auxiliaire, puisqu'il porte la flexion verbale va contrôler la distribution. Voir d'autres arguments dans la section 9.9 *Tête d'un énoncé*.

b) Nous avons vu dans la section ?? que le dépendant peut contrôler son gouverneur. Le choix de l'auxiliaire est un cas particulier de contrôle.

Corrigé 8 L'infinitif bloque la réalisation du sujet du verbe. D'après le critère d'indépendance des codépendants (section 3.3.16), un élément ne peut pas bloquer la réalisation d'un codépendant. Le syntaxème d'infinitif ne peut donc être dépendant du lexème verbal.

Corrigé 9 Une unité comme U = *le journal Libération* avec A = *le journal* et B = *Libération* met en défaut les critères distributionnels, car A et B

peuvent commuter avec U. On peut appliquer le critère topologique : en français, à l'exception de quelques adjectifs, les compléments d'un nom se placent à sa droite, ce qui laisse supposer que B dépend de A. Le critère sémantique désigne également A comme la tête : *l'affaire Dreyfus* est une affaire, tandis que *Dreyfus* est une personne.

Corrigé 10 Le test de déplacement montre l'impossibilité de permuter les syntagmes en À et en DE : *le train à Marseille de Paris, *il travaille au vendredi du lundi. Il semble donc qu'il ne s'agisse pas de deux groupes prépositionnels indépendants, lesquels peuvent généralement commuter (Elle parle de Marie à Zoé, un timbre de ma collection à 10 euros). Le test de clivage (voir la section 10.11 sur Les tests de constituance) confirme cela : *c'est au vendredi qu'il travaille du lundi vs *c'est du lundi qu'il travaille au vendredi vs c'est du lundi au vendredi qu'il travaille. Nous verrons dans le Chapitre ?? qu'on peut traiter cette construction comme une coordination.

Corrigé 11 a. On remarque que le clitique y commute avec un groupe prépositionnel comme à ça qui forme un syntagme avec sensible (sensible à ça). On a donc particulièrement \leftarrow sensible $\rightarrow y$.

b. Si l'on traite les déterminants comme dépendant du nom, le groupe substantival sujet s'analyse $un \leftarrow tiers \rightarrow de \rightarrow participants \rightarrow les$. Le reste s'analyse : $seront \rightarrow \'elimin\'es \rightarrow d\`es \rightarrow que \rightarrow commencera \rightarrow jeu \rightarrow le$

- c. La négation du français est constituée de deux mots qui se place avant et après une forme finie. C'est un argument essentiellement topologique qui nous fait dépendre pas de l'auxiliaire : $ne \leftarrow a \rightarrow pas$. On note en effet qu'on peut insérer une unité après, mais plus difficilement avant pas : elle n'a pas, comme tu sais, voulu démarrer vs ?*elle n'a, comme tu sais, pas voulu démarrer.
- d. La difficulté concerne ici le syntagme deux minutes. Celui dépend de après puisqu'il forme un syntagme avec lui (deux minutes après) et que la distribution du tout est celle de après (il est arrivé après vs *il est arrivé deux minutes).
- e. On remarque qu'avant peut former un syntagme avec moins (moins qu'avant, mais pas avec blanche (*blanche qu'avant). On a donc : blanche

→ moins → que → avant. On notera également que l'adjectif blanche est en position d'attribut, car il s'accorde avec le sujet ma chemise et non avec l'air. Il devient du coup difficile de décider si blanche dépend de a ou de air. Ici un argument topologique, l'impossibilité de permuter l'air et blanche (*elle a blanche l'air) peut nous faire préférer la dépendance à air.

f. Le problème est ici le rattachement de *épuisée par sa course*. Malgré le fait que *épuisée* s'accorde avec *Max*, il paraît difficile de considérer qu'ils forment un syntagme. En particulier, *Max* peut commuter avec *il* qui n'accepte pas de dépendant a priori. On peut noter que cette unité peut être déplacée (*Max a préféré s'arrêter, épuisé par sa course*). Nous considérons donc qu'*épuisée par sa course* dépend du verbe.

g. La question est ici la position de seuls. Il n'est pas parfaitement clair que seuls se combine avec mes amis (Qui peut accepter une telle demande? - ?? Seuls mes amis). On notera néanmoins que mes amis doit être pronominalisé par eux et pas par ils et qu'on préfère alors l'ordre inverse (*Seuls ils peuvent accepter vs ? Seuls eux peuvent accepter vs Eux seuls peuvent accepter). Ceci semble indiquer que seul se combine bien avec le substantif sujet. Notons maintenant que la tête de ce groupe n'est pas non plus très claire : en position sujet, seuls mes amis commute avec mes amis et pas avec seuls, ce qui ferait de seuls le dépendant de mes amis. Mais dans d'autres positions, mes amis ne commute plus avec seuls mes amis : *J'invite seuls mes amis; *Je parle à seuls mes amis. Au final, si l'on privilégie, le critère distributionnel avec effacement sur le critère d'effacement simple, c'est donc plutôt seuls qui semble être la tête de la combinaison seuls mes amis.

Corrigé 12 La construction causative faire Vinf du français est très cohésive et il n'est pas certain que la question posée soit légitime : on peut considérer que le sujet rétrogradé dépend du complexe verbal et pas précisément d'un des deux verbes. En particulier, dans Zoé fait dormir les enfants, on peut difficilement considérer comme valide fait les enfants ou dormir les enfants et donc le critère d'autonomisabilité ne s'applique pas. Le critère le utile semble le critère rectionnel : la réalisation du sujet rétrogradé comme objet direct, indirect ou oblique dépend complètement de la valence du Vinf et le désigne donc comme gouverneur. En effet, le sujet retrogradé aura la fonction d'objet direct si celle-ci n'est pas occupée

(même si le verbe est transitif : Zoé a fait mangé les enfants), il sera objet indirect si le Vinf a un objet direct mais pas d'objet indirect réalisé et il sera complément oblique en PAR sinon. Le critère de recouvrement s'applique et désigne plutôt faire, puisque les pronoms clitiques ne se placent pas sur le Vinf. Mais comme les compléments du Vinf se cliticisent aussi sur faire, cela n'invalide pas le choix du Vinf comme tête. Notons tout de même que, pour un énoncé comme Zoé lui fait parler à Max, le clitique lui correspond nécessairement au sujet retrogradé et à Max au destinataire des paroles, ce qui indique une plus grande proximité du sujet retrogradé avec faire.

Corrigé 13 Ces exemples montrent que le cas du groupe substantival contrôle en partie la distribution du groupe prépositionnel. On en déduit que la préposition est une tête faible (voir la section 9.15).

Corrigé 14 Les données de l'allemand montre que le complément génitif (ici *Olafs*) peut occuper aussi bien une position claire de dépendant (*das Haus Olafs*), que la position déterminative (*Olafs Haus*). Ceci tendrait à indiquer que la position déterminative, bien qu'obligatoire, peut être occupée par un dépendant du nom et que donc le déterminant n'est pas nécessairement la tête. La contrainte sur le caractère obligatoire du déterminant serait davantage de nature topologique (voir le Chapitre 11) que syntaxique.

En anglais, les données sont différentes puisque le complément génitif est uniquement possible en position déterminative et il devient plus difficile de le traiter comme un dépendant du nom. Ajoutons à cela qu'en anglais, le complément génitif peut s'employer sans le nom : We'll see each other at Olaf's 'On se verra chez Olaf'.

10 Constituance et dépendance : Différentes représentations de la combinaison

10.1 De la dépendance à la constituance

Les arbres de dépendance sont rejetés par certains linguistes sous prétexte qu'ils lient des mots, alors que les relations mettent aussi en jeu des groupes. C'est là une interprétation très réductrice des arbres de dépendances. Nous avons déjà répété à plusieurs reprises que chaque connexion ou dépendance couvre en fait une famille de combinaisons équivalentes (voir la section 8.14 *La connexion et ses instances*). Nous allons montrer dans ce chapitre que les arbres de dépendance induisent naturellement des structures de constituants et qu'inversement, une structure de constituants peut permettre de récupérer une structure de dépendance. Les deux approches sont à notre avis complémentaires et l'analyse d'une phrase complexe nécessite généralement d'identifier des unités et de les décomposer (approche de type constituance) autant que de chercher à lier entre elles des unités déjà identifiées (approche de type dépendance).

Notons encore que nous introduisons, dans ce chapitre, **deux types d'arbres de constituants**, qu'il est parfois difficile de distinguer, car ils utilisent exactement le même formalisme de représentation. Ces deux arbres, l'arbre de constituants plat (section 10.5) et l'arbre de constituants binaire (section 10.15), sont définis selon des principes différents et la correspondance avec les structures de dépendance permet de mieux comprendre ce qui les distingue. D'autres structures, équivalentes à l'un ou l'autre de ces arbres de constituants, sont également introduites.

10.2 Arbre à la Beauzée-Gladkij

Reprenons l'exemple introduit à la section 9.31 Dominance et projections maximales :

(1) Beaucoup de gens aimeraient passer Noël en Laponie.

Voici à nouveau son arbre de dépendance :

Figure 10.1 - Arbre de dépendance

Si, dans cet arbre de dépendance, les dépendants de la racine *aimeraient* sont bien les nœuds *beaucoup* et *passer*, l'interprétation linguistique ne s'arrête pas là. Ce que *beaucoup de gens aimeraient*, ce n'est pas juste *passer*, mais bien *passer Noël en Laponie*. La structure de dépendance ne dit pas le contraire. Nous allons mettre ca en évidence.

Comme nous l'avons dit et répété dans le Chapitre 8 (voir notamment la section 8.14 La connexion et ses instances), chaque connexion représente un ensemble de combinaisons équivalentes. Il en va de même pour les dépendances, qui sont des connexions hiérarchisées. Parmi toutes les combinaisons associées à une dépendance donnée, il y a bien sûr la combinaison entre le gouverneur et son dépendant. Mais il en est une autre particulièrement privilégiée par la hiérarchie : la dépendance lie d'abord et surtout un gouverneur à la projection maximale de son dépendant (voir la section 9.31 sur Dominance et projections maximales pour la définition de la projection, ainsi que l'encadré ??). Ainsi les deux dépendances qui partent de aimeraient doivent-elles être vues avant tout comme des dépendances vers les groupes beaucoup de gens et passer Noël en Laponie :

Figure 10.2 - Lien entre un nœud et les projections de ses dépendants

Cette double interprétation du dépendant, comme mot et comme groupe, est déjà bien dégagée par Nicolas Beauzée dans l'article *Régime* de l'Encyclopédie de 1765 (voir l'encadré ?? sur l'*Historique des notions de dépendance et de tête*). Elle est aussi sous-jacente chez Lucien Tesnière, qui considère la dépendance comme une relation entre mots (« dans la phrase *mon ami parle, mon* dépend de *ami*, qui dépend à son tour de *parle* »; 1959 : chap. 2), instancie les nœuds par des mots dans ses stemmas, mais définit pourtant le nœud « comme l'ensemble constitué par le régissant et par tous les subordonnés qui, à un degré quelconque directement ou indirectement, dépendent de lui et qu'il noue ainsi en quelque sorte en un seul faisceau » (Tesnières1959).

Check reference to chapters in book. We currently have 'ch', 'chap', 'chapitre'; and there are colons and commas found before.

Cette ambivalence du dépendant trouve sa première interprétation graphique avec la représentation proposée par Aleksej Gladkij en 1968, où les projections sont explicites :

Figure 10.3 - Arbre de Beauzée-Gladkij

Cette représentation, que nous appellerons désormais un Arbre de Beauzée-Gladkij, n'est pas un arbre au sens mathématique du terme, mais un arbre à bulles (voir l'encadré ?? Graphe à bulles et polygraphe): les dépendants ne sont plus des nœuds élémentaires, mais des ensembles de nœuds, représentés par des bulles. Si l'on compare l'arbre de Beauzée-Gladkij avec l'arbre de dépendance, on voit que chaque nœud a été remplacé par sa projection maximale. Laquelle projection est elle-même analysée, avec un nœud tête dont dépendent des projections, et ainsi de suite.

On aura noté que l'arbre de Beauzée-Gladkij est absolument équivalent à l'arbre de dépendance. Il ne contient aucune information supplémentaire, il ne fait qu'expliciter les projections maximales, qui étaient déjà implicitement présentes dans l'arbre de dépendance.

10.3 Arbre de constituants plat

L'arbre de Gladkij induit lui-même une nouvelle structure. Oublions les dépendances et gardons uniquement les projections : on obtient alors uniquement un **emboîtement** d'unités. Il est d'usage d'appeler les unités dans une telle structure les constituants (voir l'encadré ?? sur *Le terme* constituant).

Lorsque les constituants correspondent aux **projections maximales** des têtes comme ici, nous les appellerons les CONSTITUANTS MAJEURS.

Figure 10.4 – Emboîtement (de constituants majeurs)

(Il existe une représentation différente de la structure par un emboîtement, appelée **boîtes de Hockett**, que nous présentons dans l'encadré ?? sur l'*Historique des représentations par des diagrammes en constituants*.)

Un tel emboîtement peut être représenté sous la forme d'un parenthésage :

Notons néanmoins que le parenthésage ne peut être défini que sur une suite ordonnée d'éléments. Bien que nous ayons représenté l'emboîtement ci-dessus

```
((beaucoup (de (gens))) aimeraient (passer (Noël) (en (Laponie))))
            Figure 10.5 – Parenthésage (de constituants majeurs)
```

en remettant les mots dans l'ordre de la phrase, nous ne nous intéressons pas à l'ordre linéaire pour l'instant et un emboîtement de constituants peut être considéré indépendamment de l'ordre des mots.

Notre structure de constituants peut encore être représentée d'une autre façon : sous la forme d'un arbre (voir la section ?? Graphe et arbre), que nous appellerons l'Arbre de constituants plat, car il ne contient que les constituants majeurs et il est donc le plus plat des arbres de constituants que l'on peut considérer. Nous les représentons selon l'usage avec les mots dans l'ordre linéaire, mais nous ne nous intéressons pour l'instant qu'à la structure d'arbre, c'est-à-dire la structure hiérarchique définie par l'arbre (l'ordre sera considéré au chapitre suivant, voir notamment l'encadré ?? sur Arbre de constituants ordonné).

Figure 10.6 - Arbre de constituants plat

C'est le mathématicien Yves Lecerf, qui a le premier construit un tel arbre, immédiatement après la publication de l'ouvrage de Tesnière en 1959, et qui a établi l'équivalence entre les arbres de dépendance et cette sous-famille d'arbres de constituants. Il est facile de voir que l'arbre de constituants plats est équivalent à l'arbre de Beauzée-Gladkij : chaque nœud étiqueté d'un représente une boîte, dont les constituants sont les fils dans l'arbre. Les branches de cet arbre sont donc à lire comme des relations partie-tout et non plus comme des relations de dépendance.

10.4 Le terme constituant



Le terme constituant a été introduit par Leonard Bloomfield en 1933 dans son ouvrage Language. Bloomfield s'intéresse à la décomposition des unités et il utilise le terme pour désigner les constituants immédiats d'une unité (voir la section 8.25 Analyse en constituants immédiats). Au départ, le terme est donc prédicatif : un constituant est toujours le constituant immédiat d'une autre unité. Ensuite, le terme a été utilisé pour désigner l'ensemble des unités obtenues par l'analyse en constituants immédiats (voir la section 8.25 éponyme). Le terme n'est plus prédicatif: un constituant est devenu un type d'unité. C'est dans ce dernier emploi que nous utilisons le terme constituant. Nous considérons plusieurs types de constituants : les constituants majeurs, qui sont les projections maximales

(voir section 10.3), auxquels s'ajoutent des constituants intermédiaires obtenus par l'analyse en constituants immédiats (voir section 10.15 *Arbre de constituants binaire*).

La conséquence de ce choix terminologique est que, pour désigner les unités qui résultent de la décomposition d'un constituant C, on parle de **sous-constituants** de C.

10.5 Arbre de constituants avec têtes

Un arbre de constituants majeurs possède une propriété particulière : chaque nœud intérieur (correspondant donc à une boîte) possède parmi ses fils un unique nœud terminal — une feuille de l'arbre. (Dans notre exemple, les nœuds terminaux sont occupés par des mots, car nous avons décidé de ne pas poursuivre la décomposition au-delà. Ce choix est indépendant de la discussion présente qui reste valable quelle que soit la granularité de notre analyse.) Cette propriété d'unicité permet d'interpréter ce nœud terminal comme la tête du constituant. L'arbre de constituants plat est donc implicitement un Arbre de Constituants avec têtes.

Il est d'usage dans les arbres de constituants avec têtes (plats ou pas) d'expliciter laquelle des relations partie-tout entre ce constituant et ses sous-constituants est la relation avec la tête. Cela est indiqué par un T dans la structure suivante. Un arbre de constituant avec têtes est dit plat lorsque la tête de chaque constituant est un nœud terminal. Nous contrasterons ce type d'arbre avec les arbres X-barre et les arbres de l'analyse en constituants immédiats, qui peuvent être enrichis de têtes, mais sont binaires et non pas plats (voir la section 10.15 sur l'*Arbre de constituants binaire*). Nous présentons à cette occasion une autre convention pour l'encodage de la tête dans l'encadré ?? sur la *Syntaxe X-barre*.

Figure 10.7 - Arbre de constituants plat avec explicitation des têtes

10.6 Equivalence par la méthode de Lecerf

L'arbre de constituants plat est équivalent à l'arbre de Beauzée-Gladkij et donc à l'arbre de dépendance dont nous sommes partis. Pour reconstruire l'arbre de dépendance, rien n'est plus simple : il suffit d'identifier chaque nœud intérieur avec son nœud tête, comme le montre le diagramme suivant.

Figure 10.8 - Retour à l'arbre de dépendance

Au final, nous avons montré la parfaite équivalence entre un arbre de dépendance et un arbre de constituants plat. La méthode que nous avons utilisée pour passer de l'un à l'autre, dans un sens comme dans l'autre est due initialement à Lecerf. Nous appellerons cette méthode la MÉTHODE DE LECERF OU MÉTHODE PAR AGRÉGATION DES PROJECTIONS.

10.7 Construire un arbre de dépendance par décomposition

Dans le Chapitre 9, nous avons montré comment construire l'arbre de dépendance d'un énoncé en construisant d'abord une structure de connexion puis en raffinant et hiérarchisant cette structure grâce à différents critères permettant d'identifier les têtes des unités.

Une autre méthode consiste à construire d'abord un arbre de constituants. Cette méthode suppose qu'on ait repéré l'unité maximale que l'on veut analyser (voir notre discussion à ce propos dans le chapitre 6 et les biais que cela peut introduire dans l'analyse). Appelons cette unité une phrase selon l'usage. Voici la méthode.

A chaque étape, on a une unité U à décomposer. À la première étape, il s'agit de la phrase que l'on veut analyser. Pour décomposer U, on doit d'abord repérer la tête de U (les critères restent les mêmes que ceux que nous avons présentés dans le Chapitre 9). On enlève alors cette tête et on regarde ce que sont les morceaux qui restent. Autrement dit, on repère les sous-unités maximales de U privé de sa tête : ces unités sont les constituants majeurs de U. On recommence la même procédure avec les unités obtenues. Ainsi de suite, jusqu'à ce qu'on arrive à des unités indécomposables. (Pour plus de détails sur ce type de méthodes, voir le principe d'une *Décomposition récursive* dans l'encadré ?? qui suit).

Illustrons le procédé avec l'exemple précédent :

(2) Beaucoup de gens aimeraient passer Noël en Laponie.

On repère la tête de cette phrase par les critères exposés au chapitre précédent et notamment dans la section 9.9 : il s'agit de l'unique forme verbale finie, *aimeraient*. Les deux plus grandes unités syntaxiques une fois ce verbe retiré sont alors *beaucoup de gens* et *passer Noël en Laponie*. Ces unités syntaxiques peuvent être repérées grâce aux critères d'autonomisabilité illocutoire (section 8.11 *Unité*

syntaxique autonomisable) : elles forment des unités illocutoirement autonomisables dont le signifié est équivalent à celui qu'elles ont dans la phrase. Nous verrons dans les sections suivantes de nouveaux critères permettant d'identifier les constituants majeurs.

Cette première étape nous donne la décomposition suivante :

(beaucoup de gens) aimeraient (passer Noël en Laponie) Figure 10.9

On peut ensuite décomposer chacune des deux unités obtenues de la même façon. La tête de passer Noël en Laponie est la forme verbale passer. On voit que Nöel et en Laponie forment deux unités, notamment en appliquant le critère de déplacement (section 8.16 Tests pour la connexion) : passer en Laponie Noël est possible (et encore meilleur si on alourdit le deuxième constituant : passer en Laponie un agréable Noël). La tête de beaucoup de gens est beaucoup (critère rectionnel). On a donc la décomposition suivante :

((beaucoup (de gens)) aimeraient (passer (Noël) (en Laponie)) Figure 10.10

Encore une étape et nous aurons obtenu le parenthésage en constituants majeurs complet et donc l'arbre de constituants plat, à partir duquel nous savons construire l'arbre de dépendance, comme nous l'avons montré à la section précédente.

10.8 Décomposition récursive



Le procédé de décomposition que nous venons de présenter est dit RÉCURSIF. Le terme a un sens similaire à celui que peut avoir recours dans faire un recours, c'est-à-dire 'demander à ce que la procédure recommence'). La particularité d'un PROCESSUS RÉCURSIF est que, à la fin de chaque ÉTAPE, on se trouve avec une ou des situations similaires à celle qu'on avait au début de l'étape. Dans notre cas, chaque étape commence par un constituant à décomposer et se termine avec de nouveaux constituants à décomposer.

La récursivité est une propriété essentielle des langues, bien dégagée par Noam Chomsky dès ses premiers travaux (Structures syntaxiques en 1957). La langue se caractérise non seulement par une structure hiérarchique, mais par le fait que l'organisation à n'importe quel niveau de décomposition de la phrase s'apparente à l'organisation au niveau supérieur.

Cela est très net dans des langues comme l'anglais ou le français où la structure d'une proposition subordonnée est exactement la même que celle de la phrase complète. Cela est un peu moins vrai dans d'autres langues comme le coréen, où le verbe principal porte des marques spécifiques (voir section ?? Tête d'un énoncé dans une langue inconnue), ou comme l'allemand, où l'ordre des mots dans une subordonnée est différent de celui de la principale (voir la section ?? sur la Structure topologique de l'allemand). Néanmoins, même dans ces langues, la structure d'un groupe substantif reste exactement la même qu'il dépende du verbe principal ou du verbe d'une relative se trouvant elle-même dans une groupe substantif.

Il a été avancé que certaines langues pourraient ne pas avoir de récursivité, comme le pirahã, parlé dans la forêt amazonienne. Cette langue n'aurait pas de subordination et il ne pourrait y avoir une structure enchâssée dans une structure identique (voir la section ?? Lectures additionnelles).

10.9 Décomposition vs. combinaison

Nous avons deux grandes familles de méthodes pour construire la structure syntaxique.

La première famille de méthodes comprend des méthodes ascendantes, comme la méthode présentée aux chapitres 3.2 et 3.3. Ces méthodes reposent sur la combinaison des unités : les unités se combinent pour former des unités plus grandes, ce que nous représentons par des connexions entre les unités et des dépendances lorsque cette connexion est hiérarchisée. Elles seront appelées méthodes de construction d'un arbre de dépendance PAR COMBINAISON.

La deuxième famille de méthodes comprend des méthodes descendantes, comme la méthode de Lecerf développée plus haut (section 10.6) et la méthode par déréification développée plus bas (section 10.24). Ces méthodes reposent sur la décomposition des unités : les unités peuvent être décomposées en des unités plus petites, ce que nous représentons par des relations partie-tout, également appelée RELATIONS DE CONSTITUANCE. Elles seront appelées MÉTHODES de construction d'un arbre de dépendance PAR DÉCOMPOSITION.

Les deux familles de méthodes conduisent à la même analyse structurelle, bien qu'elles suggèrent des modes de représentations différents. Nous allons continuer à étudier cette question en détail. Nous verrons que les méthodes descendantes permettent d'introduire de nouveaux critères pour le calcul de la structure syntaxique, qu'on appelle traditionnellement les TESTS DE CONSTITUANCE, puisqu'ils sont supposés caractériser les constituants. Nous proposerons au final de combiner les deux approches pour tirer le meilleur de chacune.

Q

10.10 Les limites des méthodes par décomposition



Nous avons vu dans le chapitre précédent qu'on pouvait construire la structure de connexion à partir de n'importe quelle fragmentation de l'énoncé en raffinant ensuite les connexions en s'appuyant sur d'autres connexions (voir la section 8.22 De la fragmentation à la connexion). Rappelons qu'une fragmentation est une décomposition récursive de l'énoncé, laquelle s'apparente à une analyse en constituants immédiats (ACI), si ce n'est que les fragments ne sont pas nécessairement des constituants au sens où l'entendent les grammaires de constituants.

En un sens, la méthode de lecerf, qui permet le calcul d'un arbre de dépendance à partir d'un arbre de constituants avec têtes, est du même ordre : elle consiste à partir d'un arbre de constituants, qui est une fragmentation particulière, et à raffiner les connexions en utilisant le marquage de la tête. Il faut néanmoins remarquer que, contrairement à la méthode présentée au chapitre précédent pour la connexion, la méthode de Lecerf ne donne pas les résultats attendus si on ne part pas de la « bonne » fragmentation. Considérons l'exemple suivant :

(3) Léa a invité Théo.

Voici son analyse en constituants immédiats la plus courante et l'arbre de dépendance qui lui correspond, qui correspond à l'analyse en dépendance que nous défendons:

(4) Arbre de constituants binaire avec têtes et arbre de dépendance correspondant

Notons au passage que la méthode de Lecerf, qui nous permet de construire l'arbre de dépendance à partir de l'arbre de constituants avec têtes, ne s'applique pas qu'à des arbres de constituants plat. Ici, l'auxiliaire a est la tête d'un nœud qui est lui-même la tête de la racine de l'arbre : il faudra donc agréger les trois nœuds de cette branche pour obtenir l'arbre de dépendance, l'auxiliaire a héritant ainsi de deux dépendants, le sujet *Léa* et le participe passé *invité*.

Supposons maintenant que nous soyons partis d'une autre fragmentation :

Une autre fragmentation et l'arbre de dépendance correspondant

En partant de cet arbre de constituants avec têtes (il s'agit bien formellement d'une structure d'arbre de constituants avec têtes), il est impossible d'attribuer comme précédemment le sujet Léa à l'auxiliaire et l'objet Théo au verbe lexical invité, et ceci quels que soient les choix de tête que nous ferons; nécessairement, l'élément du syntagme *a invité* qui sera choisi comme tête héritera de tous les dépendants.

Cet exemple montre que la méthode de Lecerf ne peut aboutir à l'analyse en dépendance souhaitée que si l'on part d'une fragmentation particulière. Or la deuxième fragmentation que nous avons considérée est également une fragmentation compatible avec notre premier arbre de dépendance, où a et invité sont également connectés et forment donc un syntagme. Ne pas considérer cette deuxième fragmentation suppose d'avoir des critères pour privilégier la décomposition a invité Théo en $a \oplus invité$ Théo plutôt qu'en a $invité \oplus Théo$. Nous ne pensons pas que cela soit possible sans un raisonnement complexe qui passe nécessairement par la prise en compte de la notion de tête. Nous y reviendrons dans l'encadré ?? directement consacré à la question « Pas de structure de constituants sans la notion de tête? ». Rappelons que nous avons une autre méthode par décomposition qui consiste à raffiner l'arbre de fragmentation et qui aboutit au bon arbre de dépendance quelle que soit la fragmentation dont on part (voir la section 8.22 De la fragmentation à la connexion).

10.11 Les tests de constituance

Nous allons présenter un certain nombre de tests qui permettent de caractériser les constituants majeurs et qu'on appelle traditionnellement les TESTS DE CONSTITUANCE.

Rappelons que les **constituants majeurs** sont les **projections maximales**. Ce qui distingue l'approche constituant de l'approche projection est la façon de construire les unités. La construction des projections maximales suppose que l'on a identifié la structure de dépendance au préalable, alors que la construction des constituants majeurs ne présuppose pas que l'on connaisse la structure interne, ni même la tête d'un constituant pour l'identifier. Il peut y avoir des avantages à déterminer certains constituants majeurs avant de construire leur structure interne.

Les constituants majeurs qui dépendent d'un verbe peuvent généralement être clivés ou semi-clivés, donnant ainsi un premier test de constituance appelé le TEST DE CLIVAGE (le clivage en tant que tel sera étudié dans le Chapitre ??; les limites du test de clivage sont discutées dans la section ?? sur *Le statut des tests de constituance*). Voici ce que donne le clivage des trois constituants dépendant du verbe *va* dans l'exemple suivant :

(6) **Fred** va chez son neveu pour les vacances.

- → C'est Fred qui va chez son neveu pour les vacances.
- (7) Fred va **chez son neveu** pour les vacances.
 - \rightarrow C'est chez son neveu que Fred va pour les vacances.
- (8) Fred va chez son neveu pour les vacances.
 - → C'est pour les vacances que Fred va chez son neveu.

Ce test caractérise exactement les constituants majeurs. En particulier, il n'est pas possible de cliver deux constituants en même temps :

- (9) Fred va chez son neveu pour les vacances.
 - → *C'est chez son neveu pour les vacances que Fred va.

Les constituants majeurs présentent aussi de bonnes propriétés vis-à-vis de la pronominalisation, notamment les groupes substantifs et certains groupes prépositionnels qui peuvent être pronominalisés par des pronoms personnels ou des pronoms interrogatifs (voir l'encadré ?? qui suit à propos des *Limites des tests de pronominalisation*). Certains pronoms ne prennent pas de dépendants : c'est le cas des pronoms personnels, comme *il/elle*, *on* ou ça. La substitution d'une unité syntaxique par un tel pronom ne sera donc possible que si celle-ci n'a pas de dépendant :

- (10) Louis lit **son livre** de syntaxe. \rightarrow *Louis **le** lit de syntaxe.
- (11) Louis lit son livre de syntaxe. \rightarrow Louis le lit.

On en déduit qu'une unité substituable par un tel pronom est nécessairement une projection maximale. Dans ce cas, le TEST DE PRONOMINALISATION (voir la section 8.10 sur le *Test de commutation*) devient donc un test de constituance.

Voyons sur notre exemple comment fonctionnent les tests de constituance dont nous venons de parler. Une première salve de tests permet d'identifier *beaucoup de gens* comme un constituant, puisque celui-ci est substituable par un pronom personnel et par un pronom interrogatif et qu'il est clivable :

Pronom personnel: Ils aimeraient passer Noël en Laponie.

Pronom interrogatif: Qui aimerait passer Noël en Laponie? Beaucoup de gens.

Clivage: Il y a beaucoup de gens qui aimeraient passer Noël en Laponie.

Il en va de même pour *passer Noël en Laponie*, qui est pronomisable et semiclivable : **Pronom personnel**: Beaucoup de gens aimeraient ça.

Interrogation : Qu'est ce que beaucoup de gens aimeraient? Passer Noël en Laponie.

Semi-clivage: Ce que beaucoup de gens aimeraient, c'est passer Noël en Laponie.

Cette première étape nous donne la décomposition suivante :

```
( beaucoup de gens ) aimeraient ( passer Noël en Laponie )
Figure 10.11
```

On peut ensuite décomposer chacune des deux unités obtenues de la même façon. La pronominalisation et le semi-clivage permettent par exemple d'identifier *en Laponie* comme un constituant :

Pronom personnel: Passer Noël là.

Interrogatif: Passer Noël où? En Laponie.

Semi-clivage : Là où beaucoup de gens aimeraient passer Noël, c'est la Laponie

La-

De même pour *beaucoup de gens*, on identifie *beaucoup* comme la tête, notamment par la pronominalisation de *de gens* :

Pronom personnel: Il y **en** a **beaucoup** qui aimeraient passer Noël en ponie.

Nous obtenons le parenthésage suivant :

```
( ( beaucoup ( de gens ) ) aimeraient ( passer ( Noël ) ( en Laponie ) )

Figure 10.12
```

10.12 Les limites des tests de pronominalisation **Q**

Plusieurs remarques s'imposent concernant les tests de pronominalisation.

Première remarque : Si le test de pronominalisation appliqué avec des pronoms personnels ne reconnaît que des constituants majeurs, ce n'est pas le cas avec tous les pronoms ou proformes : à chaque fois, la substitution reconnaît une unité syntaxique (voir la section 8.10 sur le *Test de commutation*), mais pas nécessairement un constituant. Tel est le cas de la substitution par *celui* :

(12) Louis lit son livre de syntaxe. \rightarrow Louis lit celui de syntaxe.



Le cas de one en anglais est intéressant puisque, d'après plusieurs linguistes (voir l'encadré ?? de Lectures additionnelles), celui-ci peut se substituer à n'importe quelle projection du nom, y compris des projections discontinues :

Jane has a big black dog, and Bill has a brown one. 'Jane a un **gros chien** noir et Bill **un** (gros chien) marron.'

Notons que one s'utilise toujours avec un déterminant, ce qui pour le coup en fait véritablement un pro-nom et pas un pro-substantif. (En fait, ce qu'on appelle traditionnellement un pronom, comme il, est le substitut d'un substantif et pas d'un nom. Les véritables pro-noms, qui peuvent remplacer un nom, sont rares dans les langues où le nom se caractérise par la combinaison obligatoire avec un déterminant.)

On applique souvent le test de pronominalisation avec des pronoms interrogatifs, mais il faut savoir que les pronoms interrogatifs prennent parfois des dépendants ·

- (14) Qu'as tu vu d'intéressant? ← Tu as vu quelque chose d'intéressant.
- (15)Qui vient que je connais? ← Une personne que je connais vient.

En conséquence, la substitution par un pronom interrogatif n'assure pas que l'unité syntaxique reconnue soit une projection maximale.

Deuxième remarque : ce test ne peut pas être appliqué à toutes les projections maximales. Il fonctionne très bien pour les groupes substantivaux. Il fonctionne pour certains groupes prépositionnels :

- Fred va chez son neveu pour les vacances. \rightarrow Fred y va pour les (16)vacances.
- Fred va chez son neveu pour les vacances. \rightarrow ?? (17)

Il fonctionne pour certaines infinitives (celles qui commutent avec un groupe substantival):

- Lise parle de reprendre le piano depuis longtemps. \rightarrow Lise en parle depuis longtemps.
- (19) Lise recommence à jouer du piano depuis hier. \rightarrow ??

Le test ne fonctionnera pas bien pour des groupes adjectivaux, qui n'ont une proforme que dans la position prédicative, utilisée seulement marginalement :

(20) Cette représentation est **équivalente** à un arbre. → [?]Cette représentation l'est.

Conclusion : le test de pronominalisation n'est pas un test de constituance a priori. Il le devient a posteriori, une fois qu'on a identifié les pronoms qui ne peuvent pas avoir de dépendants. Il ne permet pas de décider quels types d'unités sont ou non des constituants majeurs avant d'avoir introduit des notions préalables comme la dépendance. Ce n'est qu'une fois que l'on a décidé quels types d'unités on veut considérer comme des constituants, que l'on peut tester chaque pronom pour voir s'il se substitue ou non à ces unités et constitue donc un test (avec des surprises comme dans le cas des pronoms interrogatifs). Ce test a néanmoins une réelle valeur pratique et pédagogique, une fois inventorié ce avec quoi se substitue chaque pronom. Mais, disons-le encore, il n'a aucune valeur théorique pour caractériser les constituants.

Prenons un dernier exemple. À la différence du français, l'anglais possède une proforme pour les verbes, à savoir le verbe do 'faire' :

(21) Tim plans to stay in China for a while. → Tim does.'Tim projette de rester en Chine un moment.' lit. Tim fait.

Comme tous les tests de substitution par une proforme, il caractérise des unités syntaxiques. Mais il ne permet pas de statuer sur le statut de cette unité en tant que projection maximale. Or ce test est souvent utilisé par les générativistes pour affirmer que le groupe verbal (voir l'encadré ?? sur *Le groupe verbal*) est un constituant majeur. C'est évidemment faire les choses à l'envers : c'est parce que ces linguistes ont décidé (peu importe si c'est à bon escient ou non) que le VP est un constituant qu'il voit dans la substitution par la proforme de un test de constituance.

10.13 Le statut des tests de constituance

Nous avons proposé une méthode pour construire l'arbre de constituants qui repose crucialement sur la notion de tête, puisque, avant de **décomposer un constituant**, nous cherchons **d'abord sa tête**. En d'autres termes, nous construisons directement un arbre de constituants avec têtes, lequel est formellement équivalent à un arbre de dépendance.



La question qu'on peut alors se poser est : est-il possible de construire un arbre de constituants sans faire référence à la notion de tête et sans utiliser les critères qui permettent d'identifier la tête d'une unité? En d'autres termes, a-t-on des tests suffisamment fiables pour identifier les constituants directement?

Un certain nombre de tests, dits **tests de constituance**, ont bien été proposés. Comme nous l'avons vu dans l'encadré précédent, certains tests de pronominalisation permettent de repérer certains types d'unités syntaxiques, notamment les groupes substantivaux, mais ils ne permettent pas, loin s'en faut, de caractériser tous les constituants. Nous avons également vu le test d'effacement et le test prosodique, qui nous ont permis de construire la structure de connexion (voir la section 8.8 sur Connexion et dépendance); ces tests caractérisent toutes les unités syntaxiques autonomisables et pas seulement les constituants. Les tests de déplacement et d'insertion peuvent également être utilisés pour caractériser certaines unités (voir section 8.16 sur les Tests pour la connexion), mais leur utilisation reste marginale.

Étudions plus en détail le cas du test de clivage (présenté à la section 10.11), qui est l'un des tests de constituance les plus puissants qui soit. Ce test est censé caractériser les constituants majeurs qui dépendent d'un verbe. En fait, ce test s'applique aussi à des constituants qui ne dépendent pas directement du verbe principal (voir le Chapitre ?? pour plus de détails), comme ici à sa sœur qui dépend de *parler*, alors que le verbe principal est a :

(22) C'est à sa sœur que Fred a l'intention d'essayer de parler en premier.

Mais surtout, le clivage ne s'applique pas à tous les types de constituants majeurs. Pour les complétives, le semi-clivage est mieux adapté :

- Fred espère que son neveu l'invitera. \rightarrow Ce que Fred espère, c'est que son neveu l'invitera.
 - → [?]*C'**est** que son neveu l'invitera **que** Fred espère.

Le test de clivage ne fonctionne pas non plus très bien pour les indéfinis, le test en il y a ... qu- pouvant le remplacer :

- Beaucoup de gens aimeraient venir. → Il y a beaucoup de gens qui (24)aimeraient venir.
 - → ??C'est beaucoup de gens qui aimeraient venir.

Enfin, le test de clivage ne permet pas de dégager d'unités syntaxiques en cas de négation:

- (25) Raoul n'a jamais **d'argent** sur lui. → *C'est d'argent **que** Raoul n'a jamais sur lui.
- (26) Raoul n'a **jamais d'argent** sur lui. → *C'est jamais d'argent que Raoul a sur lui

En conclusion, ce test, même s'il est l'un des plus précis, reste partiel. Il ne caractérise pas tous les constituants majeurs et il ne caractérise pas que des constituants majeurs dépendant du verbe principal. Comme pour le test de pronominalisation, ce test n'a **pas de valeur théorique** pour déterminer quels types d'unités syntaxiques sont des constituants. C'est encore une fois dans l'autre sens que les choses se font : c'est parce que nous avons déjà défini la structure syntaxique et que nous savons quelles unités syntaxiques sont des projections maximales que nous pouvons savoir quelles sont les contraintes sur le clivage et comment l'utiliser comme test.

Il n'y a donc **pas de tests** (ou de combinaisons de tests) qui permettent de **reconnaître exactement les constituants**, quelle que soit la définition qu'on en donne. Il n'y a certainement pas de tests qui permettent de reconnaître tous les constituants de la syntaxe X-barre (voir encadré *?? Binarité, construction et connexion*) et seulement ceux-là. Il n'y a pas non plus de tests qui permettent de reconnaître toutes les projections maximales et seulement les projections maximales. Tous les tests de constituance ont leurs limites et ne s'appliquent que partiellement.

Si nous mettons sérieusement en doute les tests de constituance comme outils théoriques (permettant de montrer que la constituance est une notion primitive), nous ne mettons pas en doute leur intérêt pratique et pédagogique. Ces tests peuvent et doivent être utilisés pour élaborer la structure syntaxique. À défaut de caractériser « les constituants », ils caractérisent certaines unités syntaxiques qu'il est toujours utile de pouvoir repérer (voir la section 10.28 Combiner les méthodes ascendante et descendante). Il est aussi intéressant de noter qu'ils sont finalement davantage en adéquation avec une analyse en dépendance qu'avec des analyses en constituance du type syntaxe X-barre, et ceci pour deux raisons. Premièrement, comme on l'a vu, certains tests caractérisent d'autres unités syntaxiques que les constituants traditionnels (par exemple la pronominalisation, avec des pronoms comme celui ou même les pronoms interrogatifs dans certains cas). Deuxièmement, certains types de constituants, et notamment les constituants intermédiaires de la syntaxe X-barre, ne vérifient quasiment aucun des tests et sont impossibles à caractériser de cette façon (ce qui met d'ailleurs en doute l'intérêt de les introduire dans la structure).



10.14 Pas de structure de constituants sans la notion de tête?

Dans l'encadré ?? qui précède, nous avons mis en doute la possibilité d'avoir des tests permettant de définir les constituants sans faire intervenir la notion de tête. Derrière cette question s'en cache une autre : quelles sont les notions primitives en syntaxe?

La domination des modèles basés sur la syntaxe de constituants dans la deuxième moitié du 20^e siècle a pu donner l'impression que la notion de constituant était une notion primitive, d'autant qu'il était clairement affirmé dès le départ que des notions autrefois considérées comme primitives, comme la notion de sujet syntaxique, pouvaient être ramenées à une configuration dans la structure de constituants (voir les encadrés 3.4.18 sur Binarité, construction et connexion et 3.4.19 sur la Syntaxe X-barre).

Nous avons déjà évoqué que les analyses en constituants traditionnelles privilégiaient la décomposition de a invité Théo en $a \oplus$ invité Théo sur la décomposition a invité \oplus Théo, et que cela ne nous semblait pas possible sans considérer dès le départ la notion de tête et la structure de dépendance qui en découle. Effectivement, aucun des tests de constituance présentés plus haut ne permet de privilégier une décomposition $a \oplus invité$ Théo. Ce sont des critères tels que la rection ou le contrôle de la valence supérieure qui permettent seulement de préférer ce découpage (voir la section 9.18 Tête interne et rection).

Cela devient encore plus net si on prend une forme verbale simple plutôt qu'une forme complexe et qu'on pousse la décomposition jusqu'aux syntaxèmes. Considérons:

(27) Léa invitera Théo.

Dans la structure défendue par la syntaxe X-barre (voir les encadrés 3.4.19 sur la Syntaxe X-barre et 3.4.20 sur Le groupe verbal), le sujet dépend de la flexion, tandis que l'objet dépend du lexème verbal (IP désigne un groupe flexionnel, angl. *inflexion phrase*):

inflection?

(28) Un arbre X-barre et l'arbre de dépendance correspondant

La structure de constituants ci-dessus et l'arbre de dépendance correspondant ne peuvent donc être obtenus qu'à condition de décomposer le syntagme invitera Théo en -era ⊕ invit- Théo (c'est-à-dire « flexion ⊕ INVITER Théo », avec d'autres

notations) à la place de la bien plus naturelle décomposition en *invitera* \oplus *Théo* (qui amènerait à avoir l'objet qui dépend de la flexion et non du lexème verbal; voir l'encadré ?? sur *Les limites des méthodes par décomposition*).

Notre propos n'est pas de rejeter l'analyse faite par la Syntaxe X-barre, mais simplement de souligner que cette analyse ne peut être faite sans considérer la notion de tête comme une notion primitive nécessaire à la définition des constituants. Pour notre part, nous définissons les constituants comme des projections et considérons la notion de tête comme primitive par rapport à la notion de constituant. Notre définition des constituants repose crucialement sur la notion de tête, puisque, à chaque étape, nous devons d'abord identifier la tête pour pouvoir ensuite identifier les constituants majeurs.

10.15 Arbre de constituants binaire

Nous avons donné une première définition d'un arbre de constituants basée sur les projections maximales. Nous avons appelé cet arbre l'arbre de constituants plat (voir section 10.5). Cet arbre est plat dans le sens où chaque nœud terminal (correspondant à un mot) est directement la tête de sa projection maximale et qu'il n'y a donc pas de constituants intermédiaires. Il s'agit du plus simple arbre de constituants avec têtes équivalent à un arbre de dépendance donné : les seuls constituants sont les projections maximales et les têtes (les mots dans nos exemples).

Un tel arbre a des avantages (la simplicité), mais il a aussi un inconvénient : il ne représente pas de manière explicite les connexions et le fait que ces **connexions** se font **deux à deux**. Cela a amené de nombreux linguistes utilisant les arbres de constituants comme mode de représentation syntaxique à préconiser d'avoir des ARBRES DE CONSTITUANTS BINAIRES (voir les encadrés 3.4.17 sur l'*Historique des représentations syntaxiques par des diagrammes en constituants* et 3.4.18 sur *Binarité*, *construction et connexion*). Un arbre est BINAIRE si chaque nœud intérieur a exactement deux fils. Voici l'arbre de constituants binaire le plus standard correspondant à notre exemple :

Figure 10.13 - Arbre de constituants binaire avec têtes

Dans un tel arbre, chaque décomposition correspond à une connexion, comme nous le montrerons plus en détail dans les sections 3.4.21 sur *Arbre de constituants binaire et polygraphe* et 3.4.22 sur *Polygraphe orienté et arbre de dépendance*. Voici le parenthésage correspondant à l'arbre ci-dessus avec les connexions indiquées en plus :

(beaucoup \oplus (de \oplus gens)) \oplus (aimeraient \oplus ((passer \oplus Noël) \oplus (en \oplus Laponie)))

Figure 10.14 – Parenthèsage (de constituants) binaire

Dans un arbre de constituants binaire, un nœud terminal (mot ou syntaxème selon la granularité de l'analyse) peut être la tête de plusieurs constituants : ainsi passer est-il à la fois la tête de passer Noël et de passer Noël en Laponie ou aimeraient est-il la tête de aimeraient passer Noël en Laponie et de la phrase entière. Les différents constituants dont un nœud terminal est la tête sont appelés ses projections. Le plus grand constituant dont un nœud terminal est la tête est sa projection maximale sont appelés des constituants entre un nœud terminal et sa projection maximale sont appelés des constituants intermédiaires.

Un arbre binaire contient ainsi davantage de « strates » qu'un arbre plat. Nous dirons qu'il est davantage stratifié et l'opération qui consiste à ajouter des constituants intermédiaires est appelée la STRATIFICATION.

10.16 Projections partielles et ordre de saillance

On peut définir les projections à partir de l'arbre de dépendance. Reprenons l'arbre de dépendance de notre exemple :

Figure 10.15 – Arbre de dépendance

Nous appelons projection partielle d'un nœud A de l'arbre toute unité syntaxique obtenue en prenant A et les nœuds dominés par **une partie des dépendants** de A.

Autrement dit, une projection partielle de A est obtenue en coupant éventuellement certaines des dépendances qui partent de A. Les constituants intermédiaires sont toujours les projections partielles. Mais toutes les projections partielles ne sont pas considérées comme des constituants (intermédiaires) par les analyses en constituants. Il est très généralement considéré que la projection partielle du verbe sans son sujet (aimeraient passer Noël en Laponie dans notre exemple) forme un constituant (le groupe verbal), mais pas la projection partielle du verbe sans son objet (beaucoup de gens aimeraient).

Pour obtenir un arbre de constituants binaire à partir d'un arbre de dépendance, il faut décider dans quel ordre les dépendants seront regroupés avec leur gouverneur. Ainsi pour obtenir l'arbre de constituants binaires plus haut, il faut regrouper aimeraient d'abord avec son objet (passer Noël en Laponie), puis avec son sujet (beaucoup de gens). De même, on regroupera passer avec un premier dépendant (le plus proche, Noël), puis avec un deuxième (en Laponie).

On peut poser une relation d'ordre sur les dépendances que nous appelons l'ORDRE DE SAILLANCE (l'élément le plus saillant est celui qui rejoint son gouverneur en premier). La relation d'ordre converse est appelée l'ORDRE D'OBLICITÉ (l'élément le plus oblique est celui qui rejoint son gouverneur en dernier).

Alors qu'un arbre de dépendance correspond naturellement à un arbre de constituants plat, la donnée d'un ordre de saillance permet de stratifier cet arbre : la donnée d'un arbre de dépendance avec en plus un ordre de saillance est équivalente à la donnée d'un arbre de constituants binaire avec têtes. Voici notre arbre de dépendance avec l'indication de l'ordre de saillance :

Figure 10.16 – Arbre de dépendance avec ordre de saillance

On déduit facilement de cet arbre de dépendance avec ordre de saillance un arbre de constituants binaire en regroupant tout simplement les dépendants dans l'ordre indiqué par l'ordre de saillance :

Figure 10.17 - Arbre de dépendance avec projections (maximales et partielles) équivalent à un arbre de constituants avec têtes binaire

L'arbre de constituants obtenu (celui correspondant aux emboîtements) n'est autre que l'arbre de constituants binaire avec têtes introduit à la section 10.15.

10.17 Dépendance et constituance se complètent



On dit parfois qu'un arbre de dépendance contient moins d'information qu'un arbre de constituants car il ne contient pas de projections partielles. Mais, comme on le voit, on peut facilement ajouter cette information par un étiquetage des dépendances. À l'inverse, un arbre de constituants ne contient structurellement pas d'information sur la tête : cette information doit aussi être ajoutée, soit par un étiquetage des relations partie-tout (notre T), soit par un étiquetage des nœuds (les X, X' et XP de la syntaxe X-barre, voir l'encadré ?? sur la *Syntaxe X-barre*). Dépendance et constituance se complètent donc.

Si les arbres de dépendance enrichis d'un ordre de saillance sont équivalents aux arbres de constituants (binaires ou non) enrichis de la notion de tête, il n'en reste pas moins qu'arbres de dépendance et arbres de constituants ne sont pas équivalents et que le choix repose donc sur leur différence : de l'ordre de saillance ou de la notion de tête, quelle notion voulons-nous encoder structurellement? La réponse ne fait aucun doute pour nous : la notion de tête, et c'est ce qui détermine notre forte préférence pour les structures de dépendance.

S'il est clair que la notion de tête est fondamentale, il n'est pas certain que l'ordre de saillance soit utile à la modélisation syntaxique et encore moins qu'il doive être encodé structurellement par une stratification de la structure. Nous pensons même que l'ordre de saillance induit par la stratification des arbres de constituants traditionnels n'est pas pertinent et que construire une structure comme celle de la section 10.15, où le sujet se combine obligatoirement en dernier avec le verbe, est contre-productif (voir notre discussion dans l'encadré ?? sur le *Traitement cognitif des connexions*). Il existe au contraire de nombreux arguments pour considérer que le sujet est le plus saillant des dépendants du verbe, puisque c'est le dépendant du verbe qui a le plus de « bonnes » propriétés (voir le Chapitre ?? sur les *Relations syntaxiques*).

10.18 Historique des représentations syntaxiques par des diagrammes en constituants

Si l'on excepte l'unique diagramme de Billroth de 1932, les premiers diagrammes présentant des analyses syntaxiques apparaissent à notre connaissance dans l'ouvrage de Frederick A. P. Barnard de 1836, *Analytic Grammar with Symbolic Illustrations*. Dans cette grammaire de l'anglais, le savant américain, alors âgé de 27 ans, professeur dans un asile pour sourds et muets et lui-même atteint de problèmes de surdité, propose un certain nombre de signes graphiques pour représenter les parties du discours. Mais ce qui nous intéresse surtout ici est qu'il utilise ces symboles pour décrire la structure complète d'une phrase. Barnard remarque qu'un syntagme peut se comporter de la même façon qu'un mot : par exemple, dans l'exemple qui suit, *in disposition* lit. 'en caractère' se comporte comme un adverbe. Ceci est indiqué dans le diagramme par une accolade hori-

zontale qui regroupe *in* et *disposition* et par le symbole associé à ce groupe, qui est le même que celui associé aux mots *very* et *never* et désigne donc un adverbe.

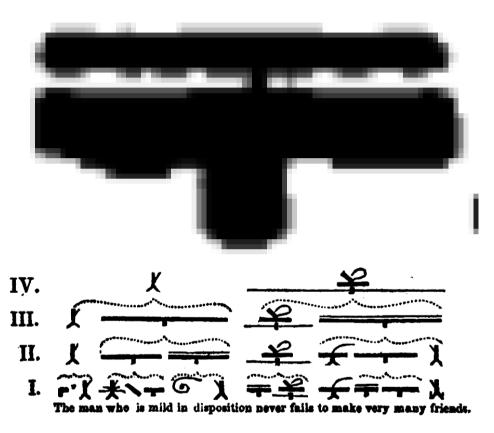


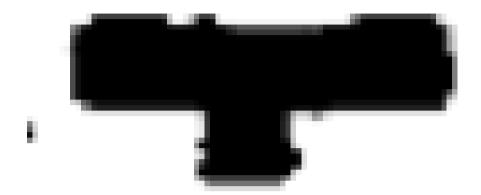
Diagramme de Barnard

L'homme qui est doux de caractère n'échoue jamais à se faire de très nombreux amis.

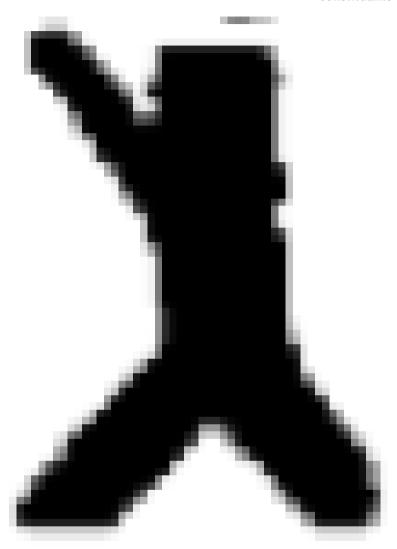
De la même façon, la relative who is mild in disposition 'qui est doux de caractère' se comporte comme un adjectif (cf. le symbole également associé à mild 'doux' et many 'nombreux') et sa combinaison avec the man 'l'homme' se comporte comme un nom (); de même, to make very many friends 'à se faire vraiment beaucoup d'amis' se comporte comme un adverbe et sa combinaison avec never fails 'n'échoue jamais' comme un verbe (). Par contre Barnard ne considère pas la proposition comme équivalente au verbe et reste avec une structure sujet-prédicat pour la proposition que l'on retrouve de la grammaire de Port-Royal à la grammaire générative (voir l'encadré ?? sur Fonctions syntaxiques et

prédicat). Le diagramme qu'il propose peut néanmoins être considéré comme un premier exemple d'arbre de constituants et même d'arbre de constituants avec tête, puisque les étiquettes indiquent la catégorie de la tête comme en syntaxe X-barre (voir l'encadré ?? sur la *Syntaxe X-barre*).





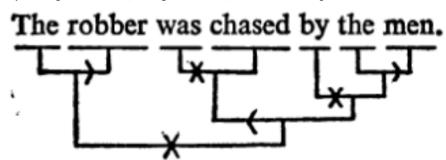
10.18 Historique des représentations syntaxiques par des diagrammes en constituants



Il semble que les contemporains de Barnard n'aient pas vraiment vu la portée de ces diagrammes et il faut attendre plus d'un siècle pour qu'en 1943, dix ans après la publication de *Langage* de Bloomfield, Eugene Nida propose à nouveau des structures de constituants, dans une thèse intitulée *A Synopsis of English Syntax*. Les structures proposées par Nida sont en fait des polygraphes (voir l'encadré ?? sur *Graphe à bulles et polygraphe*) où chaque lien représente une connexion binaire et où les liens peuvent avoir pour sommet d'autres liens. Nida

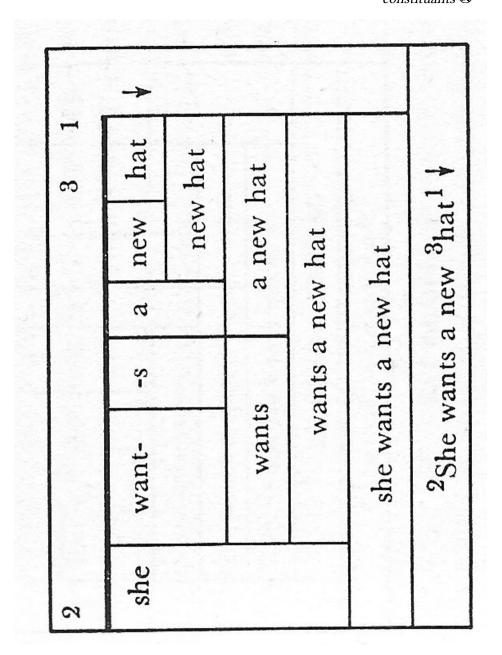
va ensuite enrichir son système en ajoutant des symboles sur les liens : les liens marquées d'une croix (X) représentent une construction exocentrique, tandis que les liens marqués d'une flèche (< ou >) représentent une construction endocentrique, la flèche pointant vers la tête. Dans le diagramme ci-dessous, *the* et *men* sont connectés par un lien > indiquant que *the* dépend de *men*, puis *by* est lié à ce lien par un lien X indiquant que la construction n'a pas de tête. Une telle représentation est équivalente à un arbre de constituants binaire (voir la section 10.22 sur *Arbre de constituants binaires et polygraphe*).

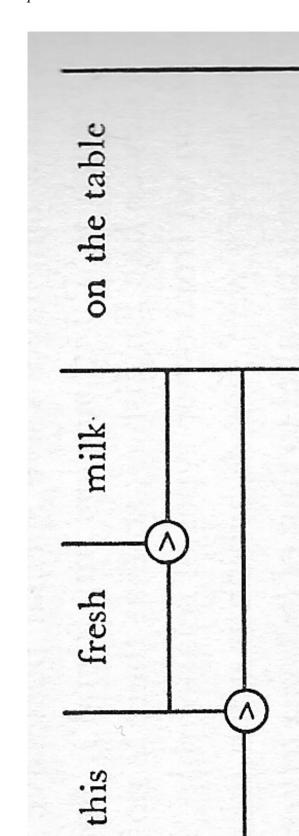
Représentation syntaxique de Nida (exemple de 1966) Le voleur a été poursuivi



par les hommes

Une représentation similaire a été proposée dans les années 1950 par Charles Hockett (voir son ouvrage A course in modern linguistics de 1958). Cette représentation, connue aujourd'hui sous le nom de Boîtes de Hockett n'est pas un arbre (au sens où les branches ne sont pas explicites), mais plutôt un emboîtement de constituants, assez proche de l'arbre de dépendance avec projections intermédiaires que nous avons présenté à la section 10.16 (sur *Projections partielles et ordre de saillance*). Les unités de bases sont les syntaxèmes (on notera la décomposition de la forme verbale wants en want- \oplus [F020?]-s) et les intonèmes (les « signes » prosodiques) sont discrétisés (c'est-à-dire représentés par des symboles, ici 2, 3, 1 et \downarrow) et également intégrés à la décomposition. Il s'agit en fait d'une décomposition assez proche de la structure syntaxique X-barre qui ré-émergera 20 ans plus tard (voir l'encadré ?? sur la $Syntaxe\ X-barre$), si ce n'est que Hockett considère encore que la combinaison sujet-verbe est endocentrique. Le diagramme peut être enrichi de symboles \boxtimes ou \boxtimes indiquant, comme chez Nida, la dépendance.





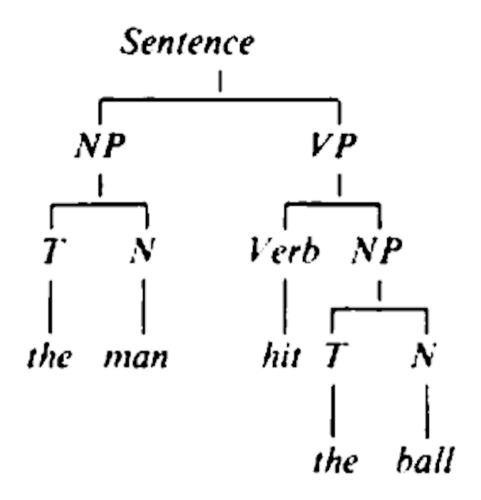
Dans cette représentation, chaque module de trois cases est à lire comme une **construction**, comme le remarque Hockett (dans le chapitre *Form classes and constructions*) :

(29) Construction chez Hockett Construction dans un arbre de constituants avec tête

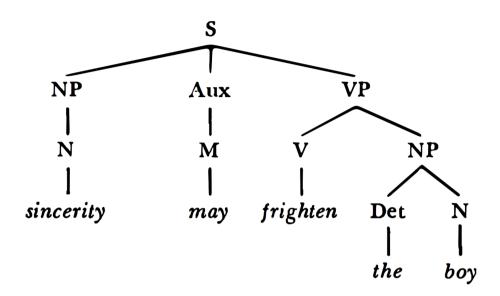
Une telle analyse est contemporaine des grammaires de réécriture de Chomsky (voir l'encadré ?? sur la *Grammaire de réécriture*) et son objectif est absolument similaire. On notera néanmoins que la représentation de Hockett met autant en avant la combinaison de A et B que la relation partie-tout de A ou de B avec le tout C qu'ils forment.

On attribue généralement l'introduction des arbres de constituants à Chomsky. On en trouve dans son article de 1956, mais il est intéressant de noter que le diagramme en constituants introduit dans Syntactic Structures de 1957 n'est pas stricto sensu un arbre. Ce diagramme, reproduit ci-dessous, apparaît dans le chapitre 4 intitulé Phrase structure 'Structure syntagmatique', où Chomsky introduit les règles de réécriture, comme $Sentence \rightarrow NP + VP$. Dans le diagramme, il n'y a pas, comme dans les arbres de constituants ordinaires, de relations partie-tout entre le nœud Sentence et les constituants immédiats NP et VP. Au lieu de cela, il y a un lien entre NP et VP correspondant au symbole + de la règle et que l'on peut interpréter comme un lien de connexion, ainsi qu'un lien entre le nœud Sentence et le lien de connexion correspondant à l'opération de réécriture représentée par le symbole \rightarrow dans la règle. Le diagramme est donc un polygraphe.

(30) Chomsky1957



(31) Chomsky1965



Dans son deuxième livre, *Aspects of the Theory of Syntax*, publié en 1965, Chomsky revient à des arbres (voir l'exemple ci-dessus pour la phrase *Sincerity may frighten the boy* 'La sincérité peut effrayer le garçon'). On notera que si le diagramme de 1957 est fondamentalement binaire, la binarité n'est plus exigée dans les arbres de 1965. Elle réapparaîtra dans les travaux suivants et notamment la syntaxe X-barre.

10.19 Binarité, construction et connexion



L'exigence de Binarité des arbres de constituants est antérieure à l'usage de la représentation en arbre. L'idée qu'une décomposition représente une Construction élémentaire est déjà bien présente dans Language de Leonard Bloomfield1933. Bloomfield donne l'exemple de la construction actor-action, que l'on pourrait appeler la construction subjectale, pour reprendre un terme proposé par Igor Mel'čuk (??). Autrement dit, il existe dans la phrase ordinaire une combinaison remarquable qui est celle du sujet avec le prédicat verbal (qu'on le conçoive comme le verbe seul ou comme un « groupe verbal », cela ne change rien pour nous, en termes de connexion). Cette construction (il s'agit bien d'une construction, au sens où on l'entend encore aujourd'hui, par exemple dans les Grammaires de constructions) est bien binaire, mettant en relation deux éléments, le sujet et le prédicat verbal. Selon les formalismes, cette construction sera représentée par une connexion ou une décomposition :

Q

(32) Trois représentations de la construction subjectale

Le premier diagramme dit qu'il y a une unité dont la tête est un nom (N) qui dépend d'un verbe (V) par une relation *sujet*, le deuxième diagramme (Chomsky1957) dit qu'il y a un groupe nominal (NP, *noun phrase*) et un groupe verbal (VP, *verbal phrase*) qui se connecte pour former ensemble une phrase (S), tandis que le troisième diagramme dit que la phrase possède deux composantes, un NP et un VP. Même si cela n'apparaît pas au premier regard, il s'agit de trois représentations de la même construction. Qui plus est, ces trois descriptions sont moins éloignées qu'il n'y paraît à première vue comme on l'a déjà vu dans le Chapitre 8 sur la *Connexion* et comme on le verra à nouveau dans la section 10.22 sur *Arbre de constituants binaires et polygraphes*, puisqu'on passe de l'une à l'autre en réifiant ou déréifiant les relations partie-tout de la relation *sujet*.

L'exigence de binarité est une propriété partagée par de nombreuses analyses en constituants et par les analyses en dépendance : de la même façon que les constructions sont encodées par des **dépendances**, c'est-à-dire des **relations deux** à **deux des mots**, dans un arbre de dépendance, les constructions seront encodées par des **décompositions binaires**, c'est-à-dire des **relations deux** à **deux des constituants**, dans un arbre de constituants binaire.

La principale différence dans les deux approches concerne la façon dont les différentes constructions sont repérées dans la structure. Dans le cadre de la syntaxe de dépendance, les constructions sont distinguées par l'étiquetage fonctionnel des dépendances (voir le Chapitre ?? sur les *Fonctions syntaxiques*) et donc par une dénomination explicite de la construction : ainsi l'étiquette *sujet* sur la dépendance plus haut indique-t-elle qu'il s'agit de la construction subjectale. Dans le cadre de la syntaxe de constituants, les primitives sont les catégories et le repérage entre les constructions est fait par la configuration et les catégories : ainsi le sujet est-il le NP sous S, tandis que l'objet est le NP sous VP :

(33) Deux représentations des constructions subjectale et objectale

10.20 Syntaxe X-barre **Q**

La Syntaxe X-barre est la version la plus aboutie de l'analyse en constituants immédiats (ACI) (voir la section 8.25 éponyme). Elle est développée à partir des années 1970 par Noam Chomsky et son étudiant Ray Jackendoff.

L'innovation majeure de la Syntaxe X-barre est que **tout constituant** est considéré comme étant la **projection d'une tête**. Dans les versions précédentes de l'ACI, la phrase était décomposée en un groupe substantival et un groupe verbal

(cf. la fameuse règle $S \rightarrow NP VP$), ce qui en faisait un constituant exocentrique. Le fait que tout constituant ait une tête rend la Syntaxe X-barre extrêmement proche d'une analyse en dépendance. (Seule la coordination est encore parfois traitée comme une construction symétrique dont les conjoints sont les co-têtes; voir encadré ?? sur les *Deux types d'arbres de constituants.*)

La deuxième caractéristique de la Syntaxe X-barre est que les nœuds de base sont des syntaxèmes et non des mots. Il y a donc deux types de constituants considérés, les projections lexicales, dont la tête est un lexème, et les projections dites fonctionnelles dont la tête est un syntaxème flexionnel (une catégorie fonctionnelle dans les termes de la Syntaxe X-barre).

La troisième caractéristique de la Syntaxe X-barre est que l'arbre de constituant est **ordonné** et aucun constituant n'est discontinu, ce qui, pour les raisons que nous avons exposé dans l'encadré ?? De la non-séparation des ordres au mouvement, entraîne la présence de nombreuses positions occupées par des traces co-indicés avec d'autres positions.

La quatrième caractéristique de l'arbre X-barre est qu'il est binaire. Ainsi la combinaison de chaque dépendant avec son gouverneur correspond à une décomposition binaire différente, ce qui, d'une certaine façon, rapproche encore plus cet arbre d'un arbre de dépendance (voir l'encadré ??). La conséquence est que si un syntaxème a beaucoup de dépendants, il sera la tête d'autant de projections. Si le syntaxème est de catégorie X, sa projection maximale est notée XP (X Phrase) et ses projections intermédiaires X' (à l'origine elles étaient notées avec une ou plusieurs barres au-dessus de X, d'où le nom de syntaxe X-barre).

La Syntaxe X-barre considère de plus que l'élément sous XP a des propriétés différentes des éléments sous X': le premier est un spécifieur, alors que les autres sont des compléments. Dans certaines versions de la Syntaxe X-barre (Jackendoff1977), les ajouts (ou modifieurs) sont structurellement distingués des compléments actanciels par l'ajout d'un niveau de stratification supplémentaire dans la configuration de base (c'est-à-dire que XP porte une « barre » supplémentaire).

Configuration de base de la Syntaxe X-barre (34)

En conclusion, l'arbre de constituants de la Syntaxe X-barre est assez différent des arbres de constituants plats. Les deux sont des arbres de constituants avec têtes, mais le dernier ne contient que des projections maximales, qui, qui plus est, peuvent former des constituants discontinus. L'arbre X-barre ne repose pas vraiment sur des tests de constituance (qui, comme on l'a vu dans les encadrés 3.4.11-12, caractérisent essentiellement les projections maximales), mais sur des **principes configurationnels**. En d'autres termes, plus la géométrie d'un arbre permet de prédire de propriétés syntaxiques de l'énoncé qu'il représente, plus l'arbre a de raisons d'être. Ainsi, par exemple, le fait que le groupe substantival qui suit *mange* possède des propriétés différentes dans les deux phrases suivantes devrait conduire à des configurations différentes, ce qui amène à compliquer à dessein la structure :

- (35) Pierre mange le pain.
- (36) Pierre mange la nuit.

Dans l'approche que nous développons ici, ces deux énoncés auront la même « structure » syntaxique (c'est-à-dire le même squelette structurel), mais les groupes substantivaux *le pain* et *la nuit* auront des fonctions totalement différentes (voir Chapitre ?? sur les *Fonctions syntaxiques*).

On peut s'étonner du fait que la syntaxe X-barre dont l'objectif est de rendre compte des propriétés syntaxiques par des configurations géométriques n'ait pas cherché à encoder structurellement une notion aussi fondamentale que la notion de *tête*, ce que font pourtant les structures de dépendance!

10.21 Le groupe verbal **Q**

Le terme groupe verbal (angl. verb phrase, VP) n'est pas utilisé de manière consistance dans la littérature. Ce terme désigne selon les auteurs deux notions différentes, que nous appellerons VP1 et VP2 :

- VP1 est une projection partielle de la forme verbale sans son sujet;
- VP2 est la projection maximale du lexème verbal.

Les deux notions sont souvent confondues, alors qu'elles renvoient à des constituants différents et à des cadres théoriques différents.

VP2 est une notion théoriquement valable, mais qui suppose que l'on travaille avec la **granularité des syntaxèmes**. Il y a alors lieu de considérer que le sujet dépend plutôt de la flexion (voir la section 8.18 *Structures de connexion, granularité et critères*) et donc VP2 est constitué du lexème verbal et de tous les éléments qu'il régit à l'exception du sujet. On notera néanmoins que VP2 n'est pas une unité autonomisable (puisque le lexème verbal n'apparaît jamais sans flexion) et qu'elle n'est donc qu'un objet théorique servant à expliciter, dans le cadre de l'analyse en constituants, que la réalisation du sujet davantage liée à la nature de la flexion qu'à celle du lexème. Il y a, à notre avis, des moyens plus simples et plus directs, d'indiquer le lien entre le sujet et la flexion verbale.

- (37) VP1 est une notion sans grand intérêt théorique de notre point de vue. Il s'agit d'un constituant intermédiaire (VP1 est le I' de l'analyse de droite ci-dessous), qui résulte d'une stratification qui nous semble peu justifiée. En effet, rien ne permet de considérer que la forme verbale se combine d'abord avec ses autres dépendants avant de se combiner avec son sujet et donc de donner ainsi un statut spécial au sujet.
 - Configurations syntaxiques caractérisant respectivement VP1 et VP2

Dernière remarque qui pourrait expliquer l'introduction de VP1 dans les analyses en constituants : l'anglais (qui est la langue la plus étudiée et la plus enseignée) a un **constituant topologique** du type VP1 (voir les exercices du Chapitre 11). Mais il est clair que la plupart des langues n'ont pas du tout cette configuration topologique et que, de toute façon, cela ne justifie pas l'introduction de VP1 dans la structure syntaxique au sens propre (que nous distinguons de la structure topologique).

10.22 Arbre de constituants binaire et polygraphe

Nous avons vu deux usages possibles de la représentation de la structure syntaxique par un arbre de constituants avec têtes : l'arbre de constituants plat et l'arbre de constituants binaire. (Nous confirmerons qu'il s'agit bien de deux usages différents du même formalisme dans l'encadré ?? sur *Deux types d'arbres de constituants*, où nous montrons l'interprétation différente du branchement ternaire dans les deux conventions de représentation).

Nous avons vu que l'arbre de constituants plat pouvait être déduit trivialement (c'est-à-dire par un procédé de conversion automatique pure, sans ajout d'aucune information) d'un arbre de dépendance, ce qui n'est pas le cas de l'arbre de constituants binaire, qui nécessite d'ajouter un ordre de saillance sur les dépendances. Cela pourrait laisser penser que l'arbre de constituants plat est plus proche de l'arbre de dépendance : c'est vrai si on se place du point de vue formel, mais ça ne l'est pas si on se place du point de vue théorique. Comme on l'a vu dans les encadrés 3.4.15-16, il y a derrière l'exigence de binarité, le souci de dégager les différentes constructions syntaxiques, de les isoler les unes des autres. Et ce souci est commun avec les grammaires de dépendances. Nous allons montrer cela maintenant en repartant de notre exemple et de son arbre binaire le plus standard :

Figure 10.18 – Arbre de constituants binaire (avec têtes)

On peut interpréter chaque branchement binaire comme une connexion entre deux unités. Comme nous l'avons vu à la section 8.12 Représenter les combinaisons, on peut représenter la combinaison $A \oplus B$ aussi bien par un lien entre A et B que par une bulle entourant A et B. Passer à un branchement binaire revient à **réifier** les relations partie-tout entre A et B et l'unité qu'ils forment ensemble (voir l'encadré ?? qui suit pour des compléments sur la notion de réification). Ici nous proposons de faire l'inverse, c'est-à-dire de **déréifier** les relations partie-tout qui constituent les branches d'un arbre de constituants. Comme en plus, chaque branchement binaire indique une tête (la branche T) et donne une connexion orientée, c'est-à-dire une dépendance :

Figure 10.19 – Passage d'un branchement binaire à une dépendance par déréification d'un nœud non terminal

Si l'on applique la déréfication à tous les nœuds intérieurs de l'arbre de constituants binaire, on obtient la structure suivante :

Figure 10.20 - Polygraphe orienté et ordonné

Cette structure est un Polygraphe (voir définition formelle dans l'encadré ?? sur *Graphe à bulles et polygraphe*) : ce n'est pas exactement un graphe, puisqu'un arc ne lie pas forcément deux nœuds (les nœuds sont les mots), mais il peut lier d'autres arcs entre eux ou un arc et un nœud. De plus, ce polygraphe est orienté, puisque chaque arc du graphe lie un gouverneur à un dépendant. Enfin ce polygraphe est implicitement ordonné, dans le sens où les nœuds du polygraphe sont disposés selon un ordre linéaire (celui des mots dans la phrase). De telles structures ont été proposées par le linguiste américain Eugene Nida dans sa thèse soutenue en 1943 (voir l'encadré ?? sur l'*Historique des représentations syntaxiques par des diagrammes en constituants*.)

10.23 Réification et transitivité

Pour se fixer les idées sur la réification, donnons un autre exemple, emprunté à la représentation sémantique des constructions transitives. Considérons une phrase élémentaire telle que :

(38) Marie frappe Pierre.

Il y a là ce qu'on appelle un verbe transitif, *frappe*, qui exprime une action d'un élément (*Marie*) sur un autre (*Pierre*). On dira encore que *Marie* est l'**agent**



de l'action et *Pierre* le **patient**. Il y a alors plusieurs façons de représenter graphiquement le « sens » de cette phrase :

(39)

Dans la représentation de gauche, *frappe* est directement modélisé comme une action de Marie sur Pierre : *frappe* (*Marie*, *Pierre*). Dans celle du milieu, les relations qui lient *frappe* à *Marie* et *Pierre* sont explicitées et nommées *agent* et *patient*. Dans celles de droite, les relations agent et patient sont vues comme des objets à part entière reliant les mots de la phrase : agent (*Marie*, *frappe*) + patient (*Pierre*, *frappe*). On l'aura deviné, les trois représentations sont des réifications successives des relations d'une même structure.

Si toute structure peut-être réifiée à l'infini, comment choisir le bon niveau de réification? Faire de la relation entre deux objets un objet n'a d'intérêt que si l'on veut la considérer en tant que telle et lui attribuer des propriétés. Dans le cas de la transitivité, on peut préférer la représentation du milieu qui explicite les relations d'agent et de patient et permet donc d'en parler, par exemple pour décrire la voix passive (voir le Chapitre ??).

Dans le cas de la relation de connexion, nous pensons qu'il n'est pas nécessaire d'expliciter les relations partie-tout que la connexion entretient avec les unités qu'elle connecte.

10.24 Polygraphe orienté et arbre de dépendance

Si l'on s'abstrait complètement de l'ordre linéaire et que l'on garde uniquement le polygraphe orienté, on peut adopter la représentation suivante, où chaque arc est représenté par une ligne droite avec le gouverneur au-dessus de son dépendant :

Figure 10.21 – Polygraphe orienté (non ordonné)

Bien que cela n'apparaisse pas au premier coup d'œil, la structure ci-dessus est un extrait de la structure précédente : c'est le même polygraphe orienté que précédemment, mais sans l'ordre des mots et avec une autre convention de représentation : au lieu d'indiquer la tête de la connexion par une flèche, celle-ci est placée au-dessus de son dépendant. Or ce polygraphe orienté, qui a donc été extrait automatiquement de l'arbre de constituants binaires en mettant en évidence les connexions, est très proche d'un arbre de dépendance. Pour obtenir l'arbre de dépendance, il suffit de faire glisser chaque dépendance le long des dépendances sur lesquelles elle s'appuie :

Figure 10.22 - « Glissade »

Nous retombons finalement sur notre arbre de dépendance :

Figure 10.23 - Retour à l'arbre de dépendance (bis)

Nous appellerons MÉTHODE PAR DÉRÉIFICATION la méthode que nous venons de présenter, qui permet de passer d'un arbre de constituants binaire avec têtes à un arbre de dépendance.

La méthode par déréification peut s'appliquer à des arbres de constituants dont certains branchements décrivent des constructions exocentriques et n'ont donc pas de marquage de la tête.

Dans l'exemple suivant (déjà étudié dans la section 9.24 Nom ou déterminant comme tête?), le petit chien dort, si on ne décide pas qui de l'article le ou du nom chien est la tête du groupe substantival, on obtient par déréification une connexion non orientée, qu'on ne peut pas faire « glisser » comme on l'a fait précédemment pour les connexions orientées.

Arbre de constituants Structure de dépendance avec marquage partiel des têtes associée

Figure 10.24

Nous retombons sur les structures de dépendance que nous avions proposé pour les constructions exocentriques au Chapitre 9.

10.25 Comparaison des méthodes par décomposition

Nous avons présenté deux **méthodes par décomposition** pour passer d'un arbre de constituants à un arbre de dépendance : la méthode de Lecerf ou méthode par agrégation des projections d'une part et la méthode par déréification d'autre part.

Les deux méthodes sont différentes et s'appliquent à des types d'arbres de constituants différents : la méthode de Lecerf s'applique uniquement à des arbres de constituants avec têtes, puisqu'elle repose crucialement sur l'agrégation des projections d'une même tête; la méthode par déréification s'applique à n'importe quel type d'arbre de constituants, mais elle ne donne des connexions binaires que si elle est appliquée à un arbre binaire (voir l'encadré ?? pour le cas des branchements ternaires).

Néanmoins lorsque les deux méthodes sont appliquées au même arbre de dépendance binaire avec têtes, elles donnent le même arbre de dépendance (voir exercices).

10.26 Équivalences entre structures

Nous avons présenté un grand nombre de structures plus ou moins équivalentes. La figure suivante rassemble ces différentes structures.

La partie droite de la figure montre le passage d'un arbre de dépendance à un arbre de constituants plat, puis retour à l'arbre de dépendance, tandis que la partie gauche montre le passage d'un arbre de dépendance à un arbre de constituants binaire avec têtes, puis retour à l'arbre de dépendance.

Chaque flèche entre deux structures symbolise une opération élémentaire permettant de passer d'une structure à une autre. Dans la plupart des cas, l'opération ne nécessite l'ajout d'aucune information et les deux structures sont équivalentes (nous ne considérons pas l'ordre linéaire sur les mots qui est des fois présent dans la structure et d'autres fois non). C'est le cas dans toute la partie droite de la figure, où toutes les structures sont équivalentes entre elles. De la même façon, toutes les structures de la partie gauche sont équivalentes entre elles. Elles contiennent une information supplémentaire qui est la stratification. Cette stratification, qui permet de stratifier les branchements non binaires, est indiquée dans la partie basse de la figur e par l'ajout de l'ordre de saillance, lequel disparaît dans la partie haute, lorsque les « glissades » sont effectuées sur le polygraphe et qu'on revient à l'arbre de dépendance.

10.27 Deux types d'arbres de constituants



Nous avons présenté dans ce chapitre deux types d'arbres de constituants avec têtes: les arbres de constituants plats et les arbres de constituants binaires comme ceux de la syntaxe X-barre. On peut penser que tout arbre de constituants plat peut être « binarisé » en introduisant des constituants intermédiaires. En fait, la différence est plus profonde et il s'agit à notre avis de deux façons assez différentes d'utiliser le même formalisme. La différence apparaît lorsqu'on regarde les branchements ternaires.

Dans un arbre plat, les branchements ternaires sont courants. Ils apparaissent dès qu'un mot a deux dépendants. Reprenons l'exemple de passer Noël en Laponie, qui se décompose en trois morceaux, une tête (passer) et ses deux dépendants (Noël et en Laponie) (voir la section sur les Arbres de constituants avec têtes plats).

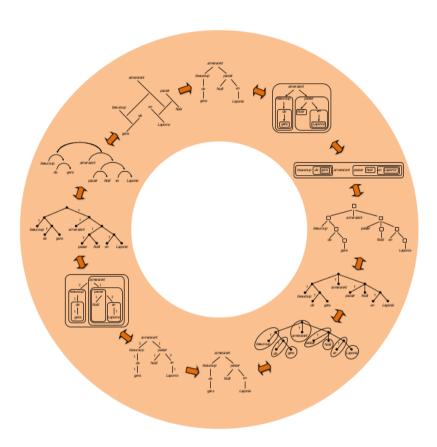


Figure 10.25



(40)

Branchement ternaire et correspondance par la méthode de Lecerf

En appliquant la méthode de Lecerf à l'arbre plat de gauche, on obtient l'arbre de dépendance à droite. Comme on le voit, le branchement ternaire est interprété comme deux connexions binaires, la connexion entre passer et Noël et la connexion entre passer et en Laponie.

Dans un arbre binaire, il en va tout autrement, puisque, comme nous l'avons montré ci-dessus (section 10.22 sur Arbres de constituants binaire et polygraphe), chaque branchement correspond à une unique connexion. Si un tel arbre contient un branchement ternaire, il doit être interprété comme une connexion ternaire (voir l'encadré ?? Connexion binaire?). La syntaxe X-barre, qui revendique l'usage des arbres binaires, s'est aventurée à proposer des branchements ternaires pour la coordination (voir Jackendoff1977, mais on trouve déjà les mêmes arbres dans Hockett1958; voir ci-dessous les boîtes de Hockett pour la coordination où le branchement est ternaire et symétrique, avec une position bien différenciée pour la conjonction).

Branchement ternaire dans une boîte de Hockett

C'est ce qu'on appelle l'analyse symétrique de la coordination (voir le Chapitre ??): dans Marie et Pierre dorment, les conjoints (c'est le nom que l'on donne aux éléments coordonnés, Marie et Pierre) sont considérés comme des co-têtes, qui contribuent à égalité à la distribution du syntagme (lequel déclenche un accord pluriel du verbe). Il s'agit alors d'une construction ternaire, que nous pouvons représenter par une connexion ternaire en suivant les conventions de représentation utilisées par Tesnière dans son article de 1934 Comment construire une syntaxe?): les deux conjoints Marie et Pierre sont au même niveau, reliés par une connexion horizontale, et la conjonction, qui est le troisième « sommet » de cette connexion, est placé en dessous.

Branchement ternaire et correspondance par la méthode de déréification

La structure de dépendance de droite peut être obtenue en appliquant la méthode de conversion par déréification à l'arbre de constituant à gauche. Le branchement ternaire donne une connexion ternaire, où Marie et Pierre occupent des positions symétriques (puisqu'ils sont co-têtes), tandis que la conjonction et occupe une troisième position, comme « dépendant » des co-têtes. La convention de Tesnière rend bien compte de cela.

En conclusion, on note que le formalisme des arbres de constituants est ambigu : il ne permet pas de distinguer une analyse plate (on ne souhaite pas considérer de constituants intermédiaires) et une connexion ternaire (on souhaite indiquer que les éléments ne se combinent pas toujours deux à deux, mais parfois par trois). Celui des structures de dépendance le permet, à condition de considérer des structures qui ne sont plus des arbres et qui contiennent des connexions ternaires.

10.28 Combiner les méthodes ascendante et descendante

A l'issu de ce chapitre et des deux précédents, nous avons présenté plusieurs méthodes pour construire la structure syntaxique. Nous aimerions conclure ce chapitre en montrant comment combiner les méthodes ascendante (par combinaison) et descendante (par décomposition). Notre présentation visait jusque-là à montrer le bien-fondé d'un certain nombre de représentations syntaxiques et leurs équivalences totales ou partielles. Mais nous n'avons pas suffisamment montré comment les différentes méthodes peuvent interagir et comment, en pratique, on peut procéder pour construire une représentation syntaxique lorsqu'on est confronté à un texte à analyser.

Nous allons prendre un exemple (on peut démarrer avec des exemples plus simples [F04A?]) (tiré des *Fleurs du mal* de Beaudelaire, *La vie antérieure*) :

- (43) C'est là que j'ai vécu dans les voluptés calmes,
- (44) Au milieu de l'azur, des vagues, des splendeurs
- (45) Et des esclaves nus, tout imprégnés d'odeurs,
- (46) Qui me rafraîchissaient le front avec des palmes,
- (47) Et dont l'unique soin était d'approfondir
- (48) Le secret douloureux qui me faisait languir.

Nous allons analyser cet exemple en procédant aussi bien par décomposition que par combinaison. Nous pouvons effectuer quatre types d'opérations.

 Nous pouvons repérer des unités syntaxiques, pas seulement des projections maximales ou partielles. Toute unité repérée peut être parenthésée ou entourée d'une bulle.

Illustration sur notre exemple : on peut par exemple s'appuyer sur la prosodie, la ponctuation ou, dans le cas d'un poème comme ici, sur le découpage en vers. Ces unités sont, en raison de leur autonomisabilité prosodique, des unités syntaxiques. En considérant que les vers et que les segments entre deux virgules sont des unités, nous obtenons le parenthésage suivant :

```
(49) [ c'est là que j'ai vécu dans les voluptés calmes ]
    [ [ au milieu de l'azur ] [ des vagues ] [ des splendeurs ] ]
    [ [ et des esclaves nus ] [ tout imprégnés d'odeurs ] ]
    [ qui me rafraîchissaient le front avec des palmes ]
    [ et dont l'unique soin était d'approfondir ]
    [ le secret douloureux qui me faisait languir ]
```

Plein d'autres découpages sont possibles. On voit en tout cas que même si notre exemple paraissait très complexe au départ, on s'est maintenant ramené à l'étude de segments beaucoup plus raisonnables.

 Nous pouvons repérer des connexions entre unités. Dès qu'une connexion est repérée, on peut tracer un arc entre deux bulles.

Notre exemple est finalement assez simple, puisque chacune des unités que nous avons dégagées à la première étape se connecte à la précédente!

```
(50) [ c'est là que j'ai vécu dans les voluptés calmes ]

-[ [ au milieu de l'azur ]-[ des vagues ]-[ des splendeurs ] ]

-[ [ et des esclaves nus ]-[ tout imprégnés d'odeurs ] ]

-[ qui me rafraîchissaient le front avec des palmes ]

-[ et dont l'unique soin était d'approfondir ]

-[ le secret douloureux qui me faisait languir ]
```

Pour vérifier que deux unités se connectent bien, il suffit de vérifier que leur combinaison donne un segment autonomisable.

 Pour tout unité syntaxique, nous pouvons repérer sa tête et, par exemple la souligner.

Cette étape est plus complexe et amène à des discussions. Il faut en particulier décider quel élément est la tête dans des groupes coordonnés comme [et des esclaves nus] ou dans une proposition relative comme [qui me rafraîchissaient le front avec des palmes]. Voir pour cela les chapitres 5.3 et 5.4. Pour d'autres unités, sans être forcément triviale, la question peut être rapidement résolue en appliquant les tests : par exemple la tête de [tout imprégnés d'odeurs] est imprégnés car tout est effaçable et d'odeurs est régi par le verbe IMPRÉGNER dont il est le complément d'agent (les odeurs imprègnent quelque chose). Nous différons donc certaines décisions et obtenons pour le début de notre exemple :

```
(51) [ c'est là que j'ai vécu dans les voluptés calmes ]

—[ [ à le milieu de l'azur ]—[ de les vagues ]—[ de les splendeurs ] ]

—[ [ et de les esclaves nus ]—[ tout imprégnés d'odeurs ] ]
```

On aura noté que certaines formes qui étaient des amalgames ont été décomposées. On vérifie en particulier que *des* contient bien la préposition DE par la commutation avec *de ces*.

 Pour toute connexion, nous pouvons repérer sa tête et la hiérarchiser pour en faire une dépendance.

La première unité contient le verbe principal de la phrase et est donc la racine de la structure, ce qui nous hiérarchise un certain nombre de connexions. Pour les autres, le critère d'effacement suffit.

```
    (52) [ c'est là que j'ai vécu dans les voluptés calmes ]
    → [ [ à le milieu de l'azur ] → [ de les vagues ] → [ de les splendeurs ] ]
    → [ [ et de les esclaves nus ] → [ tout imprégnés d'odeurs ] ]
```

Ces différentes étapes peuvent bien sûr être réalisées à tour de rôle et à plusieurs reprises :

— A chaque fois qu'une unité est décomposée en de nouvelles unités, on pourra chercher à raffiner les connexions qu'elle entretient. Quand toutes les connexions d'une unité auront été attribuées à ses sousunités, on pourra même effacer les frontières de cette unité.

Ainsi les frontières de l'unité [et des esclaves nus tout imprégnés d'odeurs] peuvent être effacées dès qu'on a établi que [et de les esclaves nus] se combinent avec [de les splendeurs].

```
    (53) [ c'est là que j'ai vécu dans les voluptés calmes ]
    → [ à le milieu de l'azur ] → [ de les vagues ] → [ de les splendeurs ]
    → [ et de les esclaves nus ] → [ tout imprégnés d'odeurs ]
    \[ qui me rafraîchissaient le front ... ]
```

 Dès qu'on a repéré la tête d'une unité, on pourra plus facilement en trouver les sous-unités (voir la section 10.7 sur Construire un arbre de dépendance par décomposition).

Par exemple, une fois repérée la tête de [tout imprégnés d'odeurs], on a immédiatement :

```
(54) tout \leftarrow imprégnés \rightarrow [ d'odeurs ]
```

Nous arrêtons là l'analyse de notre exemple. Les grands principes de l'analyse ont été donnés et nous terminons avec quelques remarques générales.

10.29 Enseigner la syntaxe

Les enseignements de syntaxe formelle se limitent en général à une partie des moyens qui précèdent. Lorsqu'on travaille en syntaxe de constituants, on demandera aux étudiants de reconnaître des projections, principalement maximales. Cela signifie que, parmi les quatre outils qui précèdent, on ne s'autorisera que l'usage du premier à savoir repérer des unités, et en plus il faudra se limiter à certaines unités seulement.

Les enseignements traditionnels en syntaxe de dépendance, eux, proposent de tracer des dépendances entre mots. Cela signifie que parmi les moyens précédents, on se limitera à tracer des connexions, uniquement entre mots, et à les orienter.

Nous proposons pour notre part de combiner tous les moyens. Toute unité syntaxique est bonne à repérer. Il n'est pas nécessaire de se limiter aux mots, comme en syntaxe de dépendance ou aux projections comme en syntaxe de constituants. Certaines séquences du type (Prép) (Dét) (Adj)* N (Adj)* (préposition-déterminantnom, avec éventuellement des adjectifs avant ou après le nom) sont très facilement repérables (voir la discussion sur les AMAS à la section ?? sur les Constituants topologiques intermédiaires). On pourra donc commencer par entourer les séquences de ce type. C'est ce que nous avons fait lorsque nous avons analysé la faible déclivité de la vallée de la Seine en Ile-de-France dans la section 9.19 sur les Tests pour la connexion. Si l'on prend notre exemple précédent, il est difficile pour un novice de repérer immédiatement le groupe substantival les esclaves nus, tout imprégnés d'odeurs, qui me rafraîchissaient le front avec des palmes, et dont l'unique soin était d'approfondir le secret douloureux qui me faisait languir, alors qu'on repèrera beaucoup plus facilement les différents amas qui constituent ce groupe : [les esclaves nus], [tout imprégnés d'odeurs], etc. C'est seulement à la fin de l'analyse, quand toutes les unités auront été connectées, que le groupe substantival en entier émergera.

Nous pensons que le moyen le plus élégant pour représenter la structure finale est d'utiliser une structure de dépendance (et nous espérons en avoir convaincu le lecteur; voir aussi l'encadré ??). Il n'y a néanmoins aucune raison de proscrire les méthodes de l'analyse en constituants, bien au contraire. Il est tout à fait possible de combiner les deux méthodes et de gagner ainsi en simplicité. L'arbre de Beauzée-Gladkij est notamment une représentation que les novices en syntaxe formelle comprennent bien et qui est moins abstraite pour eux qu'un pur arbre de constituants ou un pur arbre de dépendance. Cette représentation a en plus l'avantage de présenter simultanément les constituants et les dépendances et d'être ainsi le support idéal pour une discussion sur la distinction entre catégories et relations syntaxiques (voir chapitres 5.1 et 5.2).

Pour conclure, insistons encore une fois sur le fait qu'il n'y a pas une méthode unique pour découvrir la structure syntaxique d'un énoncé. On peut procéder aussi bien de manière descendante (décomposer une unité, notamment en repérant sa tête) comme de manière ascendante (connecter deux unités pour former une unité plus grande).



Q

10.30 Constituance ou dépendance?

Nous avons montré dans ce chapitre que les arbres de dépendance et les arbres de constituants avec têtes étaient deux modes de représentation de la structure syntaxique plus ou moins équivalents : les deux types de structures rendent compte de la façon dont les unités se combinent ou se décomposent (les connexions) et du fait que ces combinaisons sont généralement asymétriques et forment une structure hiérarchique (les têtes et les dépendances). Malgré cette équivalence formelle, les deux types de structures ont des implications théoriques différentes. Nous allons discuter ici des conséquences de cette différence en voyant les avantages et inconvénients des deux types de structures.

Têtes syntaxiques

Nous avons déjà souligné (voir la section 10.16 sur Projections partielles et ordre de saillance) que les arbres de dépendance mettent en avant la notion de tête, qui est encodée configurationnellement par les dépendances, alors que les arbres de constituants avec têtes mettent en avant l'ordre de saillance (dans leur version binaire, c'est-à-dire stratifiée). Etant donné l'importance de la notion de tête, nous pensons que c'est là un avantage fort pour les arbres de dépendances.

Constructions exocentriques

La possibilité de ne pas encoder de tête ou de considérer des co-têtes est généralement considérée comme un avantage des arbres de constituants. Nous avons vu au chapitre précédent et à nouveau ici qu'il est tout a fait possible de considérer des connexions non hiérarchisées, à côté de dépendances (c'est-à-dire des connexions hiérarchisées), dans une structure de dépendance. Cela suppose néanmoins de travailler avec des polygraphes, c'est-à-dire des « graphes » où des arêtes peuvent relier d'autres arêtes, tandis que cela peut être encodé dans un arbre de constituants sans perdre la structure d'arbre. C'est la contrepartie du fait que la notion de tête n'est pas encodée configurationnellement dans les arbres de constituants.

Unités syntaxiques

Les arbres de constituants mettent en avant certaines unités syntaxiques, à savoir les projections maximales, ainsi que des projections intermédiaires lorsque l'arbre est stratifié. Les arbres de dépendance considèrent également les projections maximales (qui correspondent aux sous-arbres de l'arbre de dépendance). Les arbres de dépendance permettent par ailleurs de considérer toutes sortes d'autres unités syntaxiques : en fait toute portion connexe de l'arbre de dépendance est une unité syntaxique potentielle (voir au chapitre précédent la définition de la *Structure de connexion*). Il ne semble pas de ce point de vue que les projections intermédiaires des grammaires de constituants soient des unités plus intéressantes que les autres, comme le montre le fait qu'elles ne vérifient en général aucun des tests de constituance. Par contre, d'autres unités, comme les nucléus (voir le Chapitre ?? sur l'*Extraction*), jouent un rôle important dans la grammaire. Or les nucléus, qui sont des chaînes immédiatement visibles dans l'arbre de dépendance, sont beaucoup plus cachés dans un arbre de constituants.

Connexions

Les connexions sont présentes dans les deux structures, mais les arbres de dépendance combinent des mots ou des syntaxèmes, tandis que les arbres de constituants combinent des constituants. Autrement dit, les instances des connexions sont plus fines dans un arbre de dépendance que dans un arbre de constituants. Ceci est pour nous un avantage majeur des arbres de dépendance.

Notons par ailleurs que si le dépendant d'une connexion est en général la projection maximale, c'est-à-dire un constituant, ce n'est pas toujours le cas. Comparons :

- (55) Les linguistes qui étaient fatigués ont quitté la conférence.
- (56) Les linguistes, qui étaient fatigués, ont quitté la conférence.

Si dans le premier exemple, le sujet du verbe doit être compris comme le constituant les linguistes qui étaient fatigués, dans le deuxième exemple, qui étaient fatigués, bien qu'étant toujours dépendant de les linguistes ne fait plus vraiment partie du sujet du verbe ont quitté. Il y a deux prédications indépendantes sur les linguistes : d'une part, les linguistes ont quitté la conférence, et d'autre part, les linguistes étaient fatigués, qui est au second plan et sert de justification à la prédication principale.



Ordre des mots

L'ordre des mots est plus facilement lisible dans un arbre de constituants, à tel point qu'il est même difficile d'envisager l'arbre de constituants indépendamment de l'ordre linéaire. Nous adoptons nous-mêmes un arbre de constituants pour représenter la structure topologique (voir chapitre suivant). Nous pensons par contre que les combinaisons syntaxiques doivent être clairement distinguées des relations de contiguïtés (voir la section 8.6 sur Structures syntaxiques et structures topologiques) et que ne pas le faire conduit immanquablement à introduire la notion de mouvement (voir la section ?? De la non-séparation des ordres au mouvement). En d'autres termes, les arbres de constituants peinent à rendre compte simplement des structures non projectives, sauf à introduire la notion de constituant discontinu, qui fait alors perdre tous les avantages de l'arbre de constituants du point de vue de l'ordre des mots. Le Chapitre 11 sera entièrement consacré à ces questions et à la définition d'une structure de constituants topologique distincte de la structure syntaxique.

Interface sémantique-syntaxe

Nous avons vu aux chapitres 2 et 6 qu'une partie de la sémantique des énoncés pouvaient être saisie par un graphe de relations prédicat-argument, lequel graphe s'apparente à une structure de connexion, c'est-à-dire à un arbre de dépendance dont on aurait retiré la hiérarchie. Ceci est très net quand on regarde des paraphrases comme Pierre a été malade pendant deux semaines vs. La maladie de Pierre a duré deux semaines (Chapitre 2). De ce point de vue, la structure sémantique est beaucoup plus proche d'un arbre de dépendance que d'un arbre de constituant.

Portée

Lorsqu'on considère un exemple tel que la première voiture rouge que j'ai vue, on remarque que l'adjectif première qui modifie voiture porte sur voiture rouge que j'ai vue et pas seulement sur voiture, dans le sens que la voiture dont on parle n'est pas première parmi toutes les voitures, mais seulement parmi les voitures rouges que j'ai vues. De tels phénomènes sont appelés des PHÉNOMÈNES DE PORTÉE. Les phénomènes de portée montrent que certaines combinaisons se font avec des groupes. On peut représenter cela par le parenthésage suivant :

```
[ la [ première [ [ voiture rouge ] que j'ai vue ] ] ] ]
```

et donc par une analyse par un arbre de constituants stratifié. De la même façon, en comparant :

- (58) les voitures coréennes chères
- (59) les voitures chères coréennes

on voit que l'interprétation est différente selon que *coréennes* est dans la portée de *chères* ou l'inverse (les voitures coréennes chères ne sont pas nécessairement des voitures chères). Ces phénomènes de portée peuvent tout de même être encodés dans un arbre de dépendance, à condition d'ajouter un ordre de combinaison avec la tête, ou par un polygraphe, en indiquant que le second adjectif se combine avec le résultat de la combinaison du premier adjectif avec le nom :

(60) les voitures coréennes chères les voitures chères coréennes Arbres de dépendance avec portée

Les structures de constituants, qui encodent la portée de manière plus simple, semblent avoir un avantage. Reste à savoir si les phénomènes de portée relèvent de la syntaxe et doivent être encodés dans la structure syntaxique ou bien s'il s'agit uniquement de phénomènes sémantiques indépendants de la structure syntaxique. Par exemple, dans une phrase telle que *Un numéro est attribué à chaque participant*, le sujet est dans la portée de l'objet, ce qui va contre l'analyse en constituants usuelle.

(61) Traitement cognitif

Nous avons déjà discuté du traitement cognitif des connexions au chapitre précédent (voir la section ?? sur le *Traitement cognitif des connexions*). Revenons rapidement sur ce point avec l'exemple suivant :

(62) J'ai rencontré un américain qui habite en Chine depuis deux ans.

Une telle phrase peut être découpée prosodiquement de la façon suivante :

(63) j'ai rencontré un américain | qui habite en Chine | depuis deux ans

Les connections entre ces groupes seront créées incrémentalement au fur et à mesure de l'écoute (ou de la lecture) :

(64) (j'ai rencontré un américain) \rightarrow (qui habite en Chine) \rightarrow (depuis deux ans)



Ceci est tout à fait compatible avec une analyse en dépendance. Une analyse en constituants immédiats suggère au contraire une analyse totalement inversée : la relative est d'abord analysée, puis rattachée à son antécédent pour former un groupe substantival, lequel est ensuite rattaché au noyau verbal pour former une proposition (sans même parler du fait que dans les analyses traditionnelles avec des arbres de constituants binaires, le sujet est censé être rattaché au verbe après le complément d'objet direct). Il est possible même avec une grammaire basée sur une analyse en constituants immédiats de faire une analyse incrémentale. Un tel algorithme a été proposé par Jay Earley en 1970 : il consiste à anticiper les constituants qui peuvent suivre un mot donné et à ouvrir de tels constituants pour les remplir ensuite par les mots qui sont analysés. Mais on voit bien qu'une telle analyse est moins naturelle qu'une analyse en dépendance.



Exercices

Exercice 1 Nous proposons le parenthésage en constituants majeurs suivant:

(65) ((son((petit) chat)) est (allergique(à(la(moquette))))) En déduire un arbre de constituant plat, puis un arbre de dépendance

en appliquant la méthode de Lecerf.

Exercice 2 Comparer:

- a. Louise va partir en Italie.
- b. Louise veut partir en Italie.

Etudiez le statut de constituant de *partir en Italie* dans ces deux exemples. Montrez que les tests donnent des résultats différents. Qu'en déduit-on sur le plan théorique?

Quel est l'intérêt du point de vue théorique d'utiliser des Exercice 3 arbres de constituants qui sont binaires?

Exercice 4 Analysez aussi finement que vous le pouvez (sans descendre en deçà des mots) la phrase suivante de Marcel Proust extraite de Du côté de chez Swan:

(66) à l'instant même où la gorgée mêlée des miettes du gâteau toucha mon palais, je tressaillis, attentif à ce qui se passait d'extraordinaire en moi.

En déduire une structure de dépendance et un arbre de constituants binaire.



Lectures additionnelles

Concernant l'analyse en constituants immédiats, on consultera les ouvrages de Bloomfiield1933, Gleason1955 et Hockett1958, plusieurs fois cités déjà, ainsi que l'article de Wells1947. Les ouvrages plus tardifs ont tendance à prendre l'ACI pour acquise et n'en discutent pas les fondements. C'est en particulier le cas des travaux de Chomsky1957; Chomsky1965. Pour l'origine des diagrammes en constituants, on consultera Barnard1936, Nida1943 et Chomsky1955; Chomsky1957. Nida1966 est une édition de sa thèse sur l'ACI soutenue en 1936, dans laquelle a été ajouté un chapitre introductif comprenant des dizaines de diagrammes syntaxiques. La syntaxe X-barre est introduite dans l'article de Chomsky1970 et l'ouvrage de Jackendoff1977. Les liens entre l'ACI et la syntaxe de dépendance sont étudiés dans un article de MazziottaKahane2017.

Les arbres de Beauzée-Gladkij ont été décrits dans l'entrée *Régime* de l'Encyclopédie et diagrammatisés par **Gladkij1968** dans un article écrit en russe.



Dans la section ?? sur la Décomposition récursive, nous avons évoqué le pirahã, une langue parlée par une tribu d'Amazonie, qui a été étudiée par le linguiste Dan Everett. Nous renvoyons à un remarquable article en ligne du New Yorker de 2007, écrit par John Colapinto, sur ce linguiste et la controverse qu'il a soulevé à propos du caractère non récursif de cette langue.

Ceux qui s'intéressent à l'algorithmique d'Earley pourront lire l'article original d'Earley1970; celui-ci est également décrit dans de nombreux ouvrages d'introduction à la théorie des langages formels.

Frederick A. P. Barnard1836 Analytic Grammar, with symbolic illustration. French, New York.

Leonard Bloomfield1933 Language, The University of Chicago Press.

Noam Chomsky1955 Three models for the description of language, IRE Transactions on information theory, 2(3), 113-124.

Noam Chomsky1957 Syntactic structures. Mouton, The Hague. [Trad. fr. M. Braudeau, Structures syntaxiques, Seuil, Paris, 1969.] Attention, dans la version française le diagramme original a été remplacé par un arbre standard!

Noam Chomsky1965 Aspects of the Theory of Syntax. The MIT Press, Cambridge. [Trad. fr. de J.-Cl. Milner, Aspects de la théorie syntaxique, Seuil, Paris, 1971.]

Noam Chomsky1970 Remarks on nominalization. In On the Nature of Grammatical Relations, Ginn and Co., Waltham, Mass, 184-221.

John Colapinto2007 Has a remote Amazonian tribe upended our understanding of language?, The New Yorker online, http://www.newyorker. com/magazine/2007/04/16/the-interpreter-2.

Jay Earley1970 An efficient context-free parsing algorithm, Communication of the ACM, 13(2), 94-102.

Gladkij Aleksej V., 1968, "On describing the syntactic structure of a sentence" (en russe avec résumé en anglais), Computational Linguistics, 7, Budapest, 21-44.

Henry A. Gleason1955 An Introduction to Descriptive Linguistics. Holt, Rinehart and Winston.

Ray **Jackendoff1977** *X* syntax : *A* study of phrase structure. MIT Press. Nicolas Mazziotta, Sylvain Kahane2017 To what extent is Immediate Constituency Analysis dependency-based? A survey of foundational texts, *Proceedings of the Fourth International Conference on Dependency Linguistics* (*Depling*), ACL, 116-126.

Eugene **Nida1943**. *Morphology : the descriptive analysis of words*. University of Michigan Press, Ann Arbor.

Eugene Nida 1966 A synopsis of English Syntax. Mouton and Co., London, The Hague.

Rulon S. Wells. 1947. Immediate constituents. Language, 23(2):81-117.





Corrections des exercices

Corrigé 1 On obtient l'arbre de dépendance suivant : (67) petit \leftarrow chat \leftarrow son \leftarrow est \rightarrow allergique \rightarrow à \rightarrow la \rightarrow moquette

Corrigé 2 Dans l'exemple b, partir en Italie peut être interrogé (Que veut Louise? - Partir en Italie.) et semi-clivé (Ce que veut Louise, c'est partir en Italie.), mais ce n'est pas possible en a (*Que va Louise?; *Ce que va Louise, c'est partir en Italie.). Dans une analyse en dépendance, nous dirons que va partir est très cohésif et qu'il n'est pas possible de les séparer, sans remettre en question que partir en Italie est une unité et que donc partir se combine avec en Italie. Dans une analyse en constituants, il n'est pas possible de considérer à la fois va partir et partir en Italie comme des unités syntaxiques, ce qui est un problème.

Corrigé 3 Un branchement binaire peut être interprété comme une connexion binaire et donc comme une construction élémentaire. Nous avons montré

dans la section 10.22-22 comment les nœuds intérieurs d'un arbre de constituants binaires pouvaient être interprétés comme des connexions, puis des dépendances en cas de marquage d'un sous-constituant tête.

Corrigé 4 En s'appuyant sur le découpage prosodique proposé par les virgules, on obtient un premier découpage : [à l'instant même où la gorgée mêlée des miettes du gâteau toucha mon palais] [je tressaillis] [attentif à ce qui se passait d'extraordinaire en moi]. On peut ensuite continuer à réduire :

```
(68) [[à l'instant même] - [[où] [[la gorgée] - [mêlée des miettes du gâteau]] - [toucha] - [mon palais]]
[[[je] - [tressaillis]] [[attentif] - [[à ce qui] - [se passait] - [d'extraordinaire] [en moi]]
```

number of braces does not match

Nous proposons finalement la structure de dépendance suivante. Nous laissons sous-spécifié la tête de la relation déterminant-nom, ainsi que la relation entre le pronom relatif et verbe principale de la proposition relative.

(69)

Pour construire un arbre de constituants binaire, il faut décider dans quel ordre combiner chaque mot avec ses dépendants. Notons que lorsqu'une connexion n'est pas hiérarchisée, cette combinaison a lieu après les autres. Par exemple, si nous prenons la portion à l'instant même où la gorgée mêlée des miettes du gâteau toucha mon palais, on obtient la structure suivante, où le déterminant est le dernier élément à se combiner avec le nom :

```
(70) [ à ( l' [ (instant même) ( où [ la (gorgée [mêlée des miettes du gâteau] ) ] [ toucha (mon palais) ] ) ] )
```

Nous devons également décider dans quel ordre regrouper les trois dépendants de la tête de la phrase, *tressaillis*. Dans ce cas, l'analyse standard regroupera probablement le sujet *je* en premier, car les deux autres dépendants sont des éléments détachés prosodiquement. Concluons en rappelant que la stratification de la structure de constituants ne nous semble



pas une question très pertinente. Nous pensons que le regroupement en constituants doit être fait à un autre niveau d'analyse en utilisant des critères de cohésion syntaxique. Ce sera l'objet du chapitre suivant.

11 La topologie : Ordre des mots, linéarisation et regroupements

11.1 La linéarité de la langue

Une des caractéristiques de la langue est la linéarité de la chaîne parlée : les sons sont prononcés les uns à la suite des autres et il en va ainsi en général des unités de la langue et notamment des mots et des syntaxèmes. Les syntaxèmes, sauf rares exceptions (voir encadré ci-dessous), se suivent selon un ordre linéaire. (Un ordre linéaire est encore dit total car tous les éléments sont ordonnés les uns par rapport aux autres.)

La PLACE qu'occupe chaque syntaxème dans l'ordre linéaire participe pleinement au sens de l'énoncé. Elle peut permettre de décider comment les unités sont combinées, et même si une combinaison est possible ou non. Par exemple, on peut envisager de placer les trois mots *Marie*, *Pierre* et *poursuit* selon 6 ordres linéaires différents :

- (1) a. Marie poursuit Pierre
 - b. Pierre poursuit Marie
 - c. Pierre Marie poursuit
 - d. Marie Pierre poursuit
 - e. poursuit Marie Pierre
 - f. poursuit Pierre Marie

Seules les combinaisons (7) et (8) portent des sens clairs en français. Et ces sens sont différents : les fonctions syntaxiques de *Marie* et *Pierre*, sujet *vs* objet du verbe POURSUIVRE, et les rôles sémantiques qu'ils expriment, agent *vs* patient du prédicat 'poursuivre', sont attribués en fonction de la place des unités dans la phrase : l'élément devant le verbe est interprété comme sujet, celui après le verbe comme objet.

Les autres combinaisons ne sont pas nécessairement dépourvues de sens (si un étranger prononçait ces énoncés, on pourrait au moins comprendre qu'il s'agit

d'une histoire de poursuite entre Pierre et Marie), mais elles ne respectent pas la syntaxe du français. Il est clair que la description précise d'une langue doit inclure les **contraintes sur l'ordre** des syntaxèmes et sur la façon dont ils se regroupent. C'est ce que nous appelons la TOPOLOGIE, du grec *topos+logos* 'étude des lieux'. Ce chapitre a comme but de proposer un modèle, appelé MODÈLE TOPOLOGIQUE ou INTERFACE SYNTAXE-TOPOLOGIE, permettant d'exprimer facilement les différentes contraintes qui existent selon les langues dans la correspondance entre la structure syntaxique et la représentation qui exprime l'ordre linéaire et les regroupements linéaires, que nous appelons la STRUCTURE TOPOLOGIQUE.

11.2 Les exceptions à la segmentabilité **Q**

Nous venons d'insister sur la linéarité de la langue et le fait que les syntaxèmes sont essentiellement prononcés les uns à la suite des autres. On appelle cette dernière propriété la SEGMENTABILITÉ: la chaîne linéaire est segmentable en une succession d'unités élémentaires (voir la section 8.6 Structure syntaxique et structure topologique).

La segmentabilité possède un certain nombre d'exceptions.

- 1. Malgré le caractère linéaire de la chaîne parlée, il est possible de jouer sur les différents paramètres du son pour communiquer deux informations simultanément. Ainsi, certains éléments de sens, comme par exemple les marqueurs d'interrogation ou de doute, ou la mise en avant de certains éléments de sens sont réalisés en utilisant la prosodie (l'intonation et le rythme la mélodie de la phrase). Ces unités les prosodèmes sont dites suprasegmentales, car elles se superposent à la chaîne des unités segmentales que sont les phonèmes. À noter que la prosodie véhicule aussi, en plus de cela, des informations utiles pour le calcul des regroupements des unités linguistiques et que les émotions du locuteur modifient également de manière significative la prosodie.
- 2. Dans les langues flexionnelles, certaines combinaisons de syntaxèmes fusionnent en un unique morphe, tant et si bien qu'on ne peut plus discerner les signifiants des différents syntaxèmes et qu'il n'y a donc plus d'ordre entre eux (voir la section 5.22 sur Amalgame, alternance et mégamorphe). Il existe aussi des syntaxèmes qui s'enchevêtrent comme dans la conjugaison des langues sémitiques, où les lexèmes verbaux sont réalisés par des consonnes et la flexion par des voyelles intercalées (voir l'encadré ?? sur les Syntaxèmes discontinus). Pour une typologie des différents cas de non-

- segmentabilité des syntaxèmes flexionnels, voir l'encadré ?? sur les *Langues isolantes*, agglutinantes et flexionnelles).
- 3. Si les syntaxèmes sont globalement continus (à l'exception des situations rappelées ci-dessus), ce n'est pas le cas des sémantèmes. Les phrasèmes peuvent être réalisés par des configurations de syntaxèmes qui ne se placeront pas nécessairement les uns à côté des autres : il ne brisa pas immédiatement la glace, « Fais pas ci, fais pas ça! ». Il faut néanmoins remarquer que la réalisation de sémantèmes discontinus passe presque uniquement par le recours à des constructions syntaxiques qui sont utilisées par ailleurs pour réaliser des combinaisons libres de syntaxèmes et qui s'appliquent aux différentes composantes du sémantème (voir Chapitre ??).
- 4. Il est important de souligner que les contraintes sur la linéarité sont l'apanage des langues vocales (c'est-à-dire utilisant la voix comme canal de communication) et de leurs contreparties écrites. Les langues des signes, réalisées par des gestes dans l'espace à trois dimensions, échappent en partie à la contrainte de linéarité. Il est possible en réalisant le signe associé à un sens donné ('chien', 'voiture', 'maison', 'maladie', etc.) de produire en même temps toutes sortes de modifieurs en adaptant la réalisation du geste : un chien agressif ou au contraire affectueux, une voiture rapide ou aux mouvements chaotiques, une grande maison, etc. Cette particularité des langues des signes par rapport aux langues vocales est généralement appelée l'ICONICITÉ: il s'agit de la possibilité de réaliser des signes iconiques, c'est-à-dire qui ont une ressemblance avec l'objet qu'il dénote. Même sans utiliser l'iconicité, la réalisation d'un signe verbal comme 'demander' (réalisé en joignant les mains à plat) indiquera en fonction de l'orientation du geste (vers le locuteur, l'interlocuteur ou un autre point de l'espace) qui demande à qui, sans qu'il soit nécessaire d'ajouter des pronoms. La diathèse du lexème est donc réalisée en même temps que le lexème. C'est bien la réalisation du signal dans un espace tridimensionnel qui permet de cumuler autant d'informations en un seul geste.

11.3 Ordre des mots vs. topologie

On renvoie traditionnellement à la branche de la syntaxe qui s'intéresse aux questions d'ordre en parlant d'Ordre des mots. Nous préférons parler de TOPOLOGIE et cela pour au moins trois raisons.

Premièrement, nous considérons que les unités minimales de la syntaxe sont les syntaxèmes et non les mots et donc que les questions d'ordre commencent

avec l'ordre des syntaxèmes à l'intérieur des mots.

Deuxièmement, nous ne nous intéressons pas seulement à l'ordre relatif des syntaxèmes (ou des mots ou des unités syntaxiques en général), mais plus généralement à la POSITION qu'occupent ces éléments dans une structure qui est plus riche qu'un simple ordre linéaire. C'est ce que désigne la topologie, qui est l'étude des lieux ou places (du grec *topos* 'lieu').

Troisièmement, le terme *topologie* (du grec *logos* 'étude'), à la différence de *ordre des mots*, renvoie clairement à un domaine d'étude de la langue; il permet de plus de dériver l'adjectif *topologique* et de parler par exemple de structure topologique, de constituant topologique ou de niveau topologique.

Le terme *topologie* provient de la tradition grammaticale allemande où le terme est d'usage depuis le milieu du 20^e siècle (voir la section ?? sur *l'Historique de la description topologique*). Les mathématiques ont également un sous-domaine nommé *topologie*, lui aussi préoccupé de la façon dont les éléments se placent les uns par rapport aux autres dans des ensembles munis d'une structure, mais sans rapport direct avec la topologie linguistique.

11.4 Place linéaire et position syntaxique

Tout unité linguistique (phonologique, morphologique, syntaxique ou sémantique) émise en contexte, dès qu'elle est segmentable, a une place dans l'ordre linéaire. Dans tout cet ouvrage, nous opposons les termes PLACE (LINÉAIRE) et POSITIONS (STRUCTURELLES) : place est toujours associé à un lieu dans l'ordre linéaire, tandis que position est associé à un lieu dans une structure plus complexe. On peut penser, comme moyen mnémotechnique, à une personne assise sur une chaise : cette personne peut changer de position sans changer de place. Les syntaxèmes sont comme des personnes assises sur des chaises les unes à la suite des autres : chacun à une place sur une chaise et peut adopter une position particulière ; il peut changer de position en restant à la même place (on parlera d'ambiguïté syntaxique) ou changer de place en gardant la même position (on parlera d'ordre libre).

Nous avons défini aux chapitres précédents la structure syntaxique. La position d'un élément dans la structure syntaxique est appelée sa position syntaxique. La syntaxe s'intéresse aux combinaisons libres (voir la section 7.6 sur *Syntaxe et morphologie*) et la position syntaxique d'un élément indique donc avec quoi il se combine, sans référence à l'ordre linéaire. Deux éléments occupent la même position syntaxique s'ils jouent un rôle similaire et s'excluent mutuellement. Ainsi dans :

- (2) Pierre regarde Marie.
- (3) Pierre la regarde.
- (4) **Qui** Pierre regarde-t-il?

les unités Marie, la et *qui* occupent la même position syntaxique, car elles correspondent toutes à la personne que Pierre regarde et elles **s'excluent mutuellement** :

- (5) *Pierre la regarde Marie.
- (6) *Qui Pierre la regarde-t-il?
- (7) *Qui Pierre regarde-t-il Marie?

Cette impossibilité pour les unités *Marie, la* et *qui* de cooccurrer implique qu'elles **occupent une même position à un certain niveau structurel**, et cela alors qu'elles n'occupent pas la même place. Ceci nous permet de conclure que la place linéaire et la position syntaxique sont bien deux notions distinctes.

11.5 Position topologique

Nous venons de voir que des éléments occupant la même position syntaxique pouvaient être à différentes places linéaires. À l'inverse des éléments de positions syntaxiques différentes peuvent venir occuper la même place linéaire.

Par exemple, en français, le sujet du verbe se place généralement devant le verbe :

(8) Des guirlandes pendaient au plafond.

Mais il est parfois possible que le sujet se place après le verbe, notamment lorsqu'un complément locatif est placé devant le verbe :

(9) Au plafond pendaient des guirlandes.

On peut ici considérer que le complément locatif vient occuper la place qu'occupe usuellement le sujet, puisqu'il n'est pas possible que le sujet se place après le verbe si le locatif ne vient pas devant le verbe :

- (10) ??Pendaient au plafond des guirlandes.
- (11) ?*Pendaient des guirlandes au plafond.

Par ailleurs, lorsque le locatif et le sujet sont tous les deux devant le verbe, le locatif est obligatoirement dans une position détachée, avec un contour prosodique particulier, sanctionné à l'écrit par une virgule :

(12) Au plafond, des guirlandes pendaient.

Lorsque le locatif est entre le sujet et le verbe, il est immédiatement interprété comme faisant partie du sujet :

(13) *Des guirlandes au plafond pendaient.

Cette exclusion mutuelle dans un lieu de l'ordre linéaire laisse supposer qu'un tel lieu est plus qu'une simple place, qu'il y a de la structure qui est en jeu. Cette structure, nous l'appellerons la structure topologique et un lieu de cette structure, comme celui où *des guirlandes* et *au plafond* commutent devant le verbe, sera appelé une position topologique. On peut ainsi reformuler ce que nous venons de dire concernant le sujet et le complément locatif : en français, il existe devant le verbe fini une position topologique qui accueille en général le sujet, mais qui peut être aussi remplie par un complément locatif, le sujet allant dans ce cas dans une autre position topologique.

11.6 Le cas des langues V2 🕏

Il existe différentes langues, dites V2, où le verbe occupe nécessairement la **deuxième place**. Autrement dit, il existe devant le verbe une unique position topologique qui doit être remplie par une et une seule unité syntaxique. Tel est le cas des langues germaniques.

Contrairement au français où toute phrase peut contenir un nombre quelconque de compléments circonstanciels antéposés au verbe, l'allemand contraint toute phrase déclarative à avoir un unique constituant devant le verbe. Et contrairement au français, ce constituant n'a pas plus à être le sujet que n'importe quel autre dépendant du verbe. Ainsi la phrase française *Malheureusement, Peter se crotte encore le nez* permet-elle des traductions à ordres variés. Parmi elles :

(14) Peter popelt leider schon wieder in seiner Nase. Peter creuse malheureusement déjà à_nouveau dans son nez.

- (15)Leider popelt Peter schon wieder in seiner Nase.
- Schon wieder popelt leider Peter in seiner Nase. (16)
- (17)In seiner Nase popelt leider schon wieder Peter.

Par contre, une traduction qui suivrait le même ordre des mots que la phrase française initiale serait agrammaticale:

(18)*Leider Peter popelt schon wieder in seiner Nase.

Ceci parce que *leider* 'malheureusement' et *Peter* ne forment pas ensemble un seul constituant. Ce constituant initial de toute phrase déclarative de l'allemand est dit occuper le CHAMP INITIAL (ou Vorfeld 'pré-champ'), le verbe fini juste après est dit être en position V2. Le champ initial (mis entre crochets dans les exemples suivants) peut être occupé par un seul constituant quelle que soit sa complexité, y compris un participe passé comme dans les exemples suivants :

seiner Nase gepopelt] hat leider wieder (19)ſΙn mal malheureusement à nouveau Peter. Dans son nez creusé Peter.

'Peter s'est malheureusement à nouveau crotté le nez.'

(20)[Vor allen mal wieder in seiner Nase zu popeln gewagt] hat Devant tous à nouveau dans son nez de creuser osé leider Peter. malheureusement Peter.

'Peter a malheureusement à nouveau osé se crotter le nez devant tout le monde.'

Comme nous l'avons montré dans la section 11.5 qui précède, le français possède aussi une forme de syntaxe V2 avec une position devant le verbe qui doit être occupée soit par le sujet, soit par un complément locatif en cas d'inversion du sujet. De ce point de vue, le français conserve des traces d'une syntaxe germanique héritée de l'époque où le latin a évolué en roman, puis en ancien français par l'intégration de peuples germaniques dans l'espace latinophone. Il s'agit d'un phénomène assez courant comparable à la créolisation : lorsqu'un groupe important de locuteurs non natifs est contraint de parler une nouvelle langue (ici le latin), il en adopte le lexique, tout en conservant partiellement la syntaxe de sa langue d'origine (ici des langues germaniques, et notamment le francique, la langue des Francs).

On retrouve des restes de ce caractère V2 en anglais, qui est aussi une langue d'origine germanique (voir l'exercice 7 sur la topologie de l'anglais). Les langues germaniques ne sont pas les seules langues V2 de par le monde : on peut mentionner par exemple le breton, une langue celtique, et le wolof, une langue Niger-Congo parlée au Sénégal. Le serbe, une langue slave, possède un phénomène comparable avec un amas de clitiques qui vient se placer en deuxième position.

11.7 Linéarisation

Il est important de souligner que nous concevons l'ordre des syntaxèmes comme le résultat d'un processus nécessaire dans l'expression du sens. Le sens n'est pas linéaire a priori (ni même hiérarchisé, voir la section 2.10 sur la *Structure hiérarchique*), mais l'usage du canal vocal oblige à un encodage linéaire de l'information. La question de l'ordre des syntaxèmes est donc vue comme un processus de LINÉARISATION de l'information.

La linéarisation est modélisée comme une correspondance entre deux structures de niveaux différents : une structure hiérarchique non ordonnée — la structure syntaxique (voir Chapitre 9) — et une structure ordonnée. Cette idée est centrale dans les *Éléments de syntaxe structurale* de Lucien **Tesnière1959**, qui distingue l'**ordre structural** et de l'**ordre linéaire**. On trouve déjà chez Claude **Buffier1709** la distinction entre syntaxe et style, puis chez Dumarsais entre syntaxe et construction. Voici ce qu'il en dit dans l'article « Construction » de l'*Encyclopédie* publié en 1754 (p. 72) :

« Je crois qu'on ne doit pas confondre *construction* avec *syntaxe*. Construction ne présente que l'idée de combinaison et d'arrangement. Cicéron a dit selon trois combinaisons différentes, *accepi litteras tuas, tuas accepi litteras*, et *litteras accepi tuas* : il y a là trois *constructions*, puisqu'il y a trois différents arrangements de mots; cependant il n'y a qu'une syntaxe; car dans chacune de ces constructions il y a les mêmes signes des rapports que les mots ont entre eux, ainsi ces rapports sont les mêmes dans chacune de ces phrases. »

Nous pensons pour notre part que la structure ordonnée est plus complexe qu'un ordre linéaire et qu'il s'agit d'un arbre de constituants ordonné que nous appelons la STRUCTURE TOPOLOGIQUE. Les CONSTITUANTS TOPOLOGIQUES jouent



notamment un rôle dans le calcul de la prosodie (voir l'encadré ?? sur Topologie et prosodie).

La grammaire qui assure la correspondance entre la structure syntaxique et la structure topologique, et donc en particulier la linéarisation, s'appelle un mo-DÈLE TOPOLOGIQUE. Dans la suite, nous allons présenter le cadre général de la construction d'un modèle topologique et nous construirons notamment les modèles topologiques du français et de l'allemand.

11.8 Une représentation syntaxique non ordonnée?



La linéarisation est une étape dans le processus de synthèse d'un énoncé à partir d'un sens (voir la section ?? La langue comme correspondance sens-son et le Chapitre 2 Production d'un énoncé). Nous pensons que l'information à communiquer par un locuteur, qui va constituer le sens de son message, n'est pas linéarisée, ni même hiérarchisée a priori. C'est le processus de communication qui oblige à linéariser. La STRUCTURE COMMUNICATIVE, qui encode la façon dont l'information est structurée pour être communiquée (voir la section ?? sur Les composantes du sens), joue ainsi dans de nombreuses langues un rôle primordial dans l'ordre des syntaxèmes (voir l'encadré ?? sur l'Ordre communicativement dirigé).

Nous considérons que la structure communicative fait partie de la représentation du sens à communiquer et ceci quelle que soit la langue, alors que l'ordre linéaire, qui en découle parfois directement, ne fait pas lui-même partie du sens. En procédant ainsi, nous pouvons nous abstraire des idiosyncrasies d'une langue particulière et nous rapprocher d'une représentation du sens plus universelle. Une telle représentation permet alors de modéliser la traduction ou le paraphrasage (qui est une traduction intra-langue) : en effet, des traductions ou paraphrases peuvent partager le même sens et des structures communicatives similaires, tout en ayant des ordres linéaires très différents (voir l'encadré ?? sur l'Ordre dominant).

Même si l'on admet que le sens est non ordonné, on est en droit de se demander pourquoi nous considérons une représentation syntaxique non ordonnée (l'arbre de dépendance) entre le sens et l'ordre linéaire.

Premièrement, on a vu qu'il est possible de définir une structure hiérarchique qui rend compte de propriétés importantes de l'énoncé sur la combinaison des unités entre elles (voir Chapitre 8) et de considérer cette structure indépendamment de l'ordre linéaire. C'est même la possibilité de complètement séparer la structure de dépendance de l'ordre linéaire qui est une des raisons de son succès en traitement automatique des langues.

Deuxièmement, il y a de bonnes raisons de penser que la structure hiérarchique est bien un intermédiaire entre le sens et l'ordre linéaire. Avant tout parce que le sens est multidimensionnel, la structure hiérarchique bidimensionnel et l'ordre linéaire unidimensionnel. Ensuite, comme nous l'avons vu au Chapitre 2, on peut associer assez naturellement une structure hiérarchique à un sens et comme nous le verrons ici on peut associer tout aussi naturellement un ordre linéaire à la structure hiérarchique. Il serait beaucoup moins simple d'associer directement un ordre linéaire à un sens.

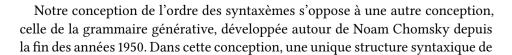
Cette dernière affirmation doit quand même être modulée : dans le processus de synthèse d'un texte à partir d'un sens, l'ordre linéaire semble parfois prévaloir sur la structure hiérarchique. On a notamment cette impression lorsqu'on compare des constructions de français oral avec des constructions plus écrites, comme dans ces exemples classiques :

- (21) Moi, mon frère, son vélo, le guidon, il est cassé.
- (22) Le guidon du vélo de mon frère est cassé.

Comme on le voit dans le premier énoncé, le plus oral des deux, chaque élément de sens forme un îlot rectionnel indépendant et cet énoncé échappe ainsi à une structure hiérarchique complète, à l'inverse du deuxième énoncé. On est en droit de se demander si dans ce premier énoncé, la structure communicative n'a pas d'abord permis de construire une structure topologique avec des places qu'on est ensuite venu remplir avec de petits segments organisés hiérarchiquement. Cela ne remet toutefois pas en cause la possibilité de considérer la structure linéaire et la structure hiérarchique indépendamment l'une de l'autre.

Dire que l'on peut considérer ces deux structures indépendamment l'une de l'autre ne signifie absolument pas qu'elles sont indépendantes l'une de l'autre. Bien au contraire, elles se contraignent l'une l'autre par un ensemble de propriétés qui constituent le module topologique. Dans la suite, nous allons donc nous intéresser au « produit » de ces deux structures, l'arbre de dépendance ordonné.

11.9 Mouvement et ordre de base





base ordonnée est postulée pour chaque construction de chaque langue. Les différents ordres observés sont obtenus par des MOUVEMENTS au sein de la structure syntaxique de base.

Le mouvement repose sur l'idée que tout changement de place linéaire est aussi un changement de position syntaxique. Nous distinguons pour notre part position topologique et position syntaxique: à chaque position syntaxique correspond au moins autant de positions topologiques qu'il y a de placements possibles. Que l'on considère le livre que Zoé lit ou le livre que lit Zoé, le nom Zoé occupe toujours la même position syntaxique de sujet du verbe lit; seule sa position topologique change. Autrement dit, à partir de la structure syntaxique commune à ces deux syntagmes, on aura deux linéarisations possibles.

Nous renvoyons également à l'encadré ?? De la non-séparation des ordres aux mouvements, où nous montrons que l'introduction du mouvement dans un modèle linguistique est une conséquence immédiate de la non-distinction de la position syntaxique et de la place linéaire, caractéristique des modèles générativistes. Dans ces modèles, il est supposé que toute construction possède un ordre de base. Ainsi en français, l'ordre de base entre le verbe et son sujet est l'ordre sujet-verbe (Zoé lit un livre). Lorsque le sujet est après le verbe (le livre que lit Zoé), on parle traditionnellement de sujet inversé. C'est une terminologie traditionnelle, mais problématique, car elle suggère qu'il y a une opération d'inversion (le livre que $Zoé lit \rightarrow le livre que lit Zoé)$ à partir de l'ordre de base.

Nous rejetons pour notre part la notion d'ordre de base. Nous considérons qu'il y a un processus de linéarisation qui permet de calculer à partir d'une structure syntaxique non ordonnée différents ordres possibles. Il est bien sûr possible qu'un des ordres soit privilégié, voire obligatoire, et nous parlerons alors d'ORDRE DOMINANT (voir les encadrés 3.5.13 sur l'Ordre dominant et 3.5.23 sur Les langues dites à ordre libre). L'ordre dominant est un ordre par défaut, généralement peu marqué communicativement, mais qui en aucun cas ne sert de base au calcul des autres ordres possibles.

L'idée d'un ordre de base et la notion d'inversion par rapport à l'ordre de base remonte au moins au 18^e siècle. L'article « Inversion » de l'*Encyclopédie* écrit par Nicolas Beauzée et publié en 1765 s'inscrit parfaitement dans son temps en faisant référence à un ordre analytique supposé l'ordre naturel de la pensée et donc universel : « C'est l'ordinaire dans toutes ces langues que le sujet précède le verbe, parce qu'il est dans l'ordre que l'esprit voie d'abord un être avant qu'il en observe la manière d'être; que le verbe soit suivi de son complément, parce que toute action doit commencer avant que d'arriver à son terme; que la préposition ait de même son complément après elle, parce qu'elle exprime de même

un sens commencé que le complément achève. » Voir également, dans l'encadré ?? Ordre communicativement dirigé, la distinction entre langues analogues et langues transpositives introduite par Girard1747. Néanmoins, à cette époque déjà, le débat est vif et d'autres linguistes ont une approche beaucoup plus nuancée, comme Dumarsais, qui dans l'article « Construction » de l'Encyclopédie publié en 1754, souligne que, s'il peut y avoir un ordre dominant, celui-ci est avant tout acquis : « La construction simple est aussi appelée construction naturelle, parce que c'est celle que nous avons apprise sans maître, par la seule constitution mécanique de nos organes, par notre attention et notre penchant à l'imitation. [...] Telle est la relation établie entre la pensée et les mots, c'est-à-dire, entre la chose et les signes qui la font connaître : connaissance acquise dès les premières années de la vie, par des actes si souvent répétés, qu'il en résulte une habitude que nous regardons comme un effet naturel. »

11.10 Arbre de dépendance ordonné

Nous avons vu au Chapitre 9 comment représenter la structure syntaxique à l'aide d'un arbre de dépendance. Nous envisageons la linéarisation comme la correspondance entre un arbre de dépendance (non ordonné) et un ordre linéaire. Si nous reprenons l'exemple des chapitres précédents (*Beaucoup de gens aimeraient passer Noël en Laponie*), nous devons mettre en correspondance les deux structures suivantes :

Figure 11.1 - Arbre de dépendance et ordre linéaire

Dans la figure précédente, l'ordre linéaire est représentée par la relation de précédence qui unit les nœuds successifs (sur le lien entre la relation d'ordre et la relation de précédence, on pourra consulter l'encadré ?? sur Dépendance, dominance et transitivité).

On peut représenter la correspondance entre les deux structures, l'arbre de dépendance et l'ordre linéaire, en alignant les nœuds qui se correspondent deux à deux :

Figure 11.2

L'ordre linéaire et l'arbre de dépendance sont en fait deux structures sur un même ensemble d'éléments. On peut donc représenter ces deux structures ensemble sur un unique ensemble de nœuds :

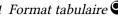


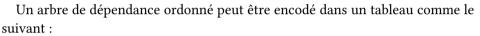
Figure 11.3

Une structure combinant ainsi deux structures est appelée une STRUCTURE PRODUIT en mathématiques. La structure obtenue par le produit d'un arbre de dépendance et d'un ordre linéaire est appelée un Arbre de dépendance ordonné.

On notera que nous avons ajouté une dépendance verticale (sans gouverneur) marquant la racine de l'arbre de dépendance. L'intérêt de cette dépendance apparaîtra clairement lorsque nous parlerons de projectivité (section 11.15 sur Projectivité et dépendances projectives).

11.11 Format tabulaire





Identifiant	Mot	Catégorie	Gouverneur	Fonction
1	Beaucoup	Adverbe	4	sujet
2	de	Préposition	1	complément
3	gens	Nom	2	complément
4	aimeraient	Verbe	0	racine
5	passer	Verbe	4	objet
6	Noël	Nom	5	objet
7	en	Préposition	5	complément
8	Laponie	Nom	7	complément

Ce tableau a 5 colonnes. La deuxième contient les mots de la phrase dans l'ordre. Pour chacun de ces mots, on peut donner autant d'informations qu'on veut : ici on donne leur catégorie syntaxique dans la troisième colonne. La première colonne attribue un identifiant à chaque mot. Grace à cet identifiant, on peut faire référence à n'importe quel mot de la même phrase. Ainsi dans la quatrième colonne, on indique pour chaque mot quel est son gouverneur. On peut ensuite ajouter des informations sur cette relation : la dernière colonne indique la fonction syntaxique que remplit chaque mot par rapport à son gouverneur.

Le tableau se lit donc ainsi : le mot 5 est passer. Ce mot est un verbe qui a pour gouverneur le mot 4 dont il est l'objet. Le mot 4 est aimeraient. Ce verbe est la racine de l'arbre de dépendance. Il n'a donc pas de gouverneur, ce que nous indiquons par le 0.

On peut représenter l'information contenue dans le tableau par une structure

étiquetée:

(23)

Ce format tabulaire, également appelé format CoNLL d'après la conférence éponyme en apprentissage automatique (*Conférence in Natural Language Learning*), est extrêmement économique et sert en traitement automatique des langues à encoder des corpus de textes analysés syntaxiquement. Les présupposés théoriques se reflètent dans l'économie du format : on ne peut encoder qu'un seul gouverneur par mot (la structure doit être un arbre) et les liens sont restreints aux mots de la même phrase. Il existe différentes extensions de ce format permettant d'encoder des graphes et mêmes des graphes à bulles.

Les corpus annotés avec des arbres de dépendance sont appelés des CORPUS ARBORÉS EN DÉPENDANCE OU BANQUES D'ARBRES DE DÉPENDANCE OU en anglais DEPENDENCY TREEBANKS. Ils servent pour des études sur les propriétés d'une langue donnée, aussi bien qu'à l'apprentissage automatique d'outils logiciels comme des analyseurs syntaxiques automatiques.

11.12 Notation polonaise inverse 🕏

Nous allons nous intéresser à la structure d'un calcul algébrique élémentaire et voir que le rapport entre une structure arborescente et l'ordre n'est pas qu'une question interne à la linguistique. La question a d'ailleurs intéressé les mathématiciens avant les linguistes.

Considérons la formule algébrique $(7 \times 19) + (5 \times 31)$. Pour effectuer ce calcul, il faut nécessairement effectuer les deux multiplications avant de les sommer, ce que l'on peut représenter par le diagramme suivant, où le résultat de chaque calcul intermédiaire est donné sous une barre horizontale :

(24)

On peut alors donner une représentation arborescente de la structure de ce calcul. Les signes + et × représentent des opérateurs binaires qui combinent deux nombres pour en fournir un troisième. Ces opérateurs ont donc deux arguments à l'image d'un verbe transitif et une représentation similaire, par un arbre de dépendance, peut être adoptée :

De telles structures de dépendance pour les formules ont été introduites à la suite des travaux du logicien polonais Jan Łukasiewicz, qui a montré vers 1920 qu'il y avait plusieurs façons d'encoder linéairement un tel arbre. Une première façon est celle que nous utilisons quand nous écrivons la formule $(7 \times 19) + (5 \times 10^{-5})$



 \times 31), où à chaque fois l'opérateur binaire a été placé entre ses deux dépendants. Cette écriture, dite infixée, nécessite des parenthèses pour ne pas être ambiguë : en effet, $7 \times 19 + 5 \times 31$ peut tout aussi bien correspondre à $7 \times (19 + (5 \times 31))$, ((7 \times 19) + 5) \times 31 ou encore (7 \times (19 + 5)) \times 31.

Une autre écriture, dite ÉCRITURE POLONAISE OU PRÉFIXÉE, consiste à lire l'arbre en ramassant toujours le gouverneur avant ses dépendants et les dépendants de gauche à droite : + × 7 19 × 5 31 (voir *Exercices* en fin de chapitre). Cette formule n'est pas ambiguë, malgré l'absence de parenthèses : il suffit de connaître l'ARITÉ de chaque symbole, c'est-à-dire le nombre d'arguments et donc de dépendants qu'il a. Les opérateurs + et × sont BINAIRES, c'est-à-dire d'arité 2, tandis que les nombres sont d'arité 0.

La lecture postfixée ou polonaise inverse, où le gouverneur est ramassé après ses dépendants, est particulièrement appropriée au calcul et utilisée par les calculateurs automatiques, y compris certaines machines à calculer d'usage courant : en effet, pour effectuer le calcul 7 $19 \times 5 31 + \times$, il suffit d'empiler les nombres au fur et à mesure de la lecture et à la lecture de chaque opérateur binaire d'effectuer l'opération sur les nombres des deux lignes qui précèdent en les supprimant :

(25)

Ces différentes stratégies pour passer d'une structure hiérarchique à un ordre linéaire se retrouvent dans les langues naturelles (voir l'encadré?? qui suit). Ceci montre la nécessité de distinguer, pour les calculs comme pour les énoncés linguistiques, la structure et son encodage linéaire. Notre calcul, encodé par la formule $(7 \times 19) + (5 \times 31)$, aussi bien que par les formules $+ \times 7.19 \times 5.31$ ou $7.19 \times 5.31 + 1.00$ ×, est totalement indépendant de la convention qui sert à l'encoder linéairement. Autrement dit la structure du calcul n'est pas ordonnée : seule la structure hiérarchique encodée par l'arbre de dépendance est pertinente. (Avec des opérateurs non commutatifs, comme la soustraction ou la division, l'ordre sur les fils d'un opérateur peut également être pertinent.) L'ordre linéaire résulte uniquement d'une convention d'encodage. Ceci est en grande partie vrai des langues naturelles : l'ordre linéaire des syntaxèmes imposé par la grammaire de la langue est une convention pour encoder la structure syntaxique de l'énoncé. Autrement dit, la structure syntaxique d'un énoncé, qui indique comment les éléments de cet énoncé se combinent les uns aux autres, est une structure qui est fondamentalement indépendante de l'ordre conventionnellement utilisé pour l'encoder linéairement.

11.13 Langues à têtes finales et langues à têtes initiales 🍑



Les travaux en typologie, notamment le travail fondateur de Joseph Greenberg (Universals of Language, 1963), puis l'étude de Matthew S. Dryer1992 sur plus de 500 langues, ont pu montrer une corrélation entre la place du verbe dans la phrase et la place du nom dans le groupe nominal ou de l'adjectif dans le groupe adjectival : ainsi les langues qui placent le verbe en fin de phrase tendent très fortement à placer les noms en fin de groupe nominal et vice versa.

Henri Weil (1818-1909) a anticipé ces résultats en classant les langues selon la place de la tête par rapport à ses dépendants. Il distingue ainsi les LANGUES DES-CENDANTES où la tête précède ses dépendants (et où en avançant dans la phrase on descend dans l'arbre et on s'éloigne de la racine) des LANGUES MONTANTES où la tête suit ses dépendants (et où en avançant dans la phrase on monte dans l'arbre et on se rapproche de la racine). Lucien Tesnière (1989-1954), en comparant près de 200 langues (comme on peut le voir dans ses archives conservées à la Bibliothèque National de France), a repris cette classification en distinguant les langues centrifuges (du latin centrum 'centre' et fugio 'fuir') où les dépendants suivent leur gouverneur des LANGUES CENTRIPÈTES (de peto 'tendre vers') ou les dépendants précèdent leur tête. On parle aujourd'hui de langues à têtes FINALES et de LANGUES À TÊTES INITIALES (et en anglais de head-final et headinitial languages).

Nous allons comparer des langues représentatives de ces deux tendances, en nous basant sur l'exemple suivant :

Les étudiants très motivés organisent une conférence internationale.

Le coréen est une langue à têtes finales, et la traduction de cette phrase en coréen donne une phrase où tous les gouverneurs se trouvent à droite de leurs dépendants :

- (27)maeu euiyokjeok-i-n haksaeng-deul-i kukje hakhoi-leu jojikha-nda très être motivé-Q étudiant-PL-NOM international conférence-ACC organiser-PRES
 - (Q : translatif en qualificatif, PL : pluriel, NOM : nominatif, ACC : accusatif, PRES: présent)



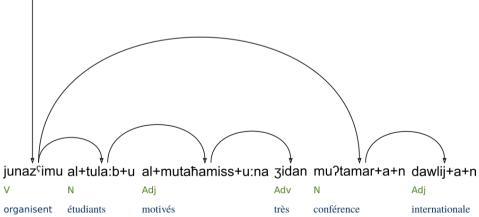


La même phrase traduite en arabe classique (l'arabe du 7^e siècle utilisé dans le Coran) possède une structure inversée : tous les gouverneurs se trouvent devant leurs dépendants, y compris le verbe qui se trouve au tout début de la phrase. L'arabe classique est une langue à têtes initiales.

(28)junaz¹imu al+tula :b+u al+mutaħamiss+u :na ʒidan mu?tamar+a+n dawlij+a+n

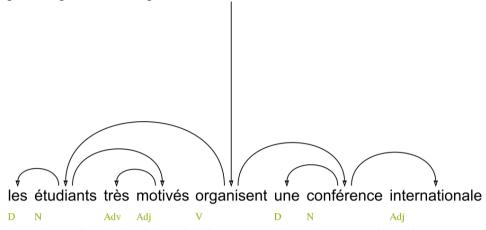
V D+N+NOM D+Adj+NOM ADV N+ACC+D Adj+ACC organise.IND.PRES les étudiants les enthousiasmés très une conférence internationale

(IND : indicatif, PRES : prés, NOM :nominatif, ACC :accusatif)



Le français est une langue intermédiaire. L'arbre de dépendance linéarisé de

la phrase 1 possède des dépendances dans les deux directions :



Notez quand même qu'en arabe classique, comme en coréen, il y a des exceptions à la règle du placement de la tête, et les notions de têtes initiales et têtes finales doivent être vue comme des tendances, plutôt que des règles absolues.

11.14 Classification des langues selon l'ordre V, S, O

Depuis les travaux de Greenberg, il est d'usage de classer les langues selon l'ordre dominant entre le verbe (V), son sujet (S) et son objet (O). On considère qu'il y a un ordre dominant lorsqu'un des six ordres possibles entre V, S et O domine significativement les autres. C'est le cas du français où l'ordre SVO domine largement. La place des S et O pronominaux n'est pas prise en compte, car celle-ci diffère parfois de l'ordre dominant, comme en français où la cliticisation de l'objet (*il la regarde*) entraine un ordre SOV.

Le World Atlas of $\mathit{Language}$ $\mathit{Structures}$ Online recense actuellement :

put this in tabular?

565 langues SOV, soit 41%

488 langues SVO, soit 35%

95 langues VSO, soit 7%

25 langues VOS, soit 1,8%

11 langues OVS, soit 0,8%

4 langues OSV, soit 0,3%

et 189 langues sans ordre dominant, soit 14%.

Comme on le voit, les langues à têtes finales sont beaucoup plus répandues que les langues à têtes initiales et le sujet a nettement tendance à se placer avant

l'objet. Il existe une grande proportion de langues SVO, comme le français, où le verbe tend à se placer entre son sujet et son objet, mais l'ordre le plus répandu est quand même l'ordre SOV, celui du coréen. Il est à noter que le classement des langues selon ce principe présuppose que les notions de sujet et d'objet soient pertinentes pour n'importe quelle langue et qu'on soit capable de les identifier. Nous verrons au Chapitre ?? que la question est loin d'être simple en raison de l'existence de langues dite **ergatives**. Dans le classement ci-dessus, c'est en fait l'agent et le patient, plus simple à identifier que le sujet et l'objet, qui correspondent à S et O (voir Chapitre ??).

Nous poursuivrons la discussion sur topologie et typologie dans l'encadré ?? sur les *Langues dites à ordre libre*, qui n'ont généralement pas d'ordre dominant.

11.15 Projectivité

Nous avons déjà évoquée la projectivité à deux reprises au travers du test d'insertion (section 8.16 *Tests pour la connexion*) et du test de recouvrement (section 9.33 éponyme).

La projectivité est une propriété qui contraint la linéarisation, c'est-à-dire la correspondance entre un arbre de dépendance et un ordre linéaire. Ce n'est ni une propriété de la structure dépendance, ni une propriété de la structure d'ordre : c'est une **propriété** de la structure produit, c'est-à-dire de l'arbre de dépendance ordonné.

Intuitivement, un arbre de dépendance dont tous les mots se placent autour de leur gouverneur est dit projectif. Donnons une première définition formelle de la projectivité :

Un arbre de dépendance ordonné est dit PROJECTIF si chacune de ses **projections** (maximales) est **continue**, c'est-à-dire forme une portion continue de la chaîne linéaire.

Le terme *projectivité*, forgé par le mathématicien Yves Lecerf en 1960, renvoie directement à la notion de projection. Dire que la projection de A est continue revient à dire que les dépendants de A se sont placés autour de A et les dépendants de ses dépendants aussi et ainsi de suite et que donc l'ensemble forme un

segment continu autour de A.

L'arbre de dépendance ordonné présenté tout à l'heure est projectif :

Figure 11.4 - Arbre de dépendance projectif

En effet, chaque projection est continue. Par exemple, la projection de *passer*, qui est l'ensemble formé par les mots *passer*, *Noël*, *en* et *Laponie*, forme le segment *passer Noël en Laponie*, qui est une portion linéairement continue.

La projectivité entraîne une certaine asymétrie des dépendances. En effet, si A gouverne B, alors la projection de B sera entièrement d'un des deux côtés de A, tandis que la projection de A pourra très bien avoir des parties à gauche et à droite de B.

Figure 11.5 - Projections maximales continues

C'est sur cette asymétrie qu'est basé le test de recouvrement (section 9.33 éponyme).

Il existe une autre caractérisation de la projectivité, qui peut s'observer immédiatement sur une représentation graphique comme celle ci-dessus, où les nœuds sont placés sur une ligne et toutes les dépendances forment des arcs placés du même côté de la ligne.

Un arbre de dépendance ordonné est projectif si et seulement si les dépendances ne se coupent pas.

Cela concerne également la dépendance verticale qui marque la racine et qui doit être vue comme une dépendance infinie empêchant tout autre dépendance de passée au-dessus de la racine. Autrement dit, les deux configurations suivantes sont exclues d'un arbre de dépendance projectif :

Figure 11.6 - Configurations non projectives

Donnons un exemple de structure non projective :

(29) Un petit cordonnier qui voulait aller danser Avait fabriqué de petits souliers.



Une belle est entrée qui voulait les acheter,

Mais le cordonnier lui a déclaré : « Ils seront à vous sans qu'il vous coûte un sou. Mais il vous faudra danser avec moi. » (Le petit cordonnier, chanson de Francis Lemarque, 1953)

Le troisième vers de cette chanson a une structure non projective :

Figure 11.7 - Arbre de dépendance non projectif

Nous avons fait ici une analyse très peu granulaire. Toute analyse plus fine, au niveau du mot par exemple, conservera ces trois dépendances et donc la configuration non projective. (Faire une analyse moins granulaire revient à réduire le graphe. Voir la définition de la réduction dans l'encadré qui suit.)

Comme on le voit sur la figure, la dépendance qui relie une belle à la relative qui voulait les acheter couvre la racine est entrée et coupe donc la dépendance verticale. Il s'en suit que le groupe substantival une belle qui voulait les acheter est discontinu.

La projectivité est en partie énoncée par Nicolas Beauzée dans l'article « Régime » de l'Encyclopédie publié en 1765 (voir la discussion dans l'encadré ?? Historique des notions de dépendance et de tête) : « Il ne faut jamais rompre l'unité d'un complément total, pour jeter entre ses parties un autre complément du même mot. », ce qui reviendrait effectivement à ce que le premier complément soit discontinu. Beauzée note également que les langues à cas peuvent violer la projectivité: « je crois qu'il est bon de remarquer, que les règles que je viens d'assigner sur l'arrangement de divers compléments, ne peuvent concerner que l'ordre analytique qu'il faut suivre quand on fait la construction d'une phrase, ou l'ordre usuel des langues analogues comme la nôtre. Car pour les langues transpositives, où la terminaison des mots sert à caractériser l'espèce de rapport auquel ils sont employés, la nécessité de marquer ce rapport par la place des mots n'existe plus au même degré. » La notion de langue analogue renvoie ici aux langues ayant un ordre fixe similaire au français (voir l'encadré ?? sur l'Ordre communicativement dirigé).

11.16 Projectivité et planarité 🕏



La deuxième définition de la projectivité que nous avons donnée (en termes de coupure) est un cas particulier d'une propriété plus générale des graphes que l'on appelle la PLANARITÉ. (Pour la notion de graphe voir l'encadré ?? sur Graphe et arbre.)

Un graphe est dit PLANAIRE s'il peut être dessiné dans un plan sans qu'aucunes de ses arêtes ne se coupent.

Dans le cas de la projectivité d'un arbre ordonné, on s'intéresse à une propriété qui concerne la relation entre la structure d'arbre et l'ordre linéaire sur les nœuds de l'arbre. Le graphe qui nous intéresse doit donc contenir ces deux structures. Au graphe que forme l'arbre, nous allons ajouter l'ordre linéaire sous la forme d'arêtes entre deux nœuds successifs, ce qui revient à placer les nœuds du graphe sur une ligne dans l'ordre linéaire. Pour prendre totalement en compte la structure hiérarchique de l'arbre, il faut encoder le fait que l'arbre possède un nœud, la racine, qui domine tous les autres. Pour cela, on attache l'arbre par sa racine à un nœud spécial qu'on appelle le nœud à l'infini (noté ∞). Cela revient à placer les nœuds de l'arbre sur un cercle dans l'ordre linéaire avec le nœud à l'infini entre le premier et le dernier nœud. Nous appellerons ce graphe le GRAPHE CIRCULAIRE associé à un arbre linéairement ordonné.

Graphe circulaire (31)

Le graphe circulaire d'un arbre projectif a une propriété un peu plus forte que la planarité : il est planaire extérieur.

Un graphe est dit planaire extérieur s'il peut être dessiné avec tous ses nœuds accessibles de l'extérieur et sans qu'aucunes de ses arêtes ne se coupent.

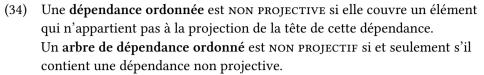
On peut montrer que pour qu'un graphe soit planaire extérieur, il faut et il suffit qu'il ne puisse pas être réduit à un graphe particulier qu'on appelle K₄. La RÉ-DUCTION d'un graphe consiste à agréger ensemble des nœuds qui sont liés. C'est une opération importante de notre point de vue, car elle consiste à regarder la structure avec une granularité moins importante (voir la section ?? sur la Granularité). Le graphe K₄ est le graphe complet à 4 nœuds, c'est-à-dire le graphe à 4 nœuds où tous les nœuds sont liés deux à deux. Les deux configurations rejetées par la projectivité sont exactement le graphe K₄, comme le montre la figure suivante.

(33) Equivalence entre K₄ et les configurations non projectives

11.17 Équivalence des définitions de la projectivité 🍳



On peut donner encore une autre caractérisation de la non-projectivité :



Soit T un arbre de dépendance ordonné. On veut montrer l'équivalence des trois propriétés suivantes :

P1 : T contient une dépendance non projective.

P2 : T contient un nœud dont la projection est discontinue.

P3 : T contient deux dépendances qui se coupent.

Supposons P1. Alors il existe un dépendance $x \to y$ non projective. Cette dépendance couvre un ancêtre z de x. Par définition, z n'est pas dominé par x et n'appartient donc pas à la projection X de x. Comme z est entre xet y qui appartiennent à X, X est donc discontinu, d'où P2. Supposons maintenant P2. Soit X un projection discontinue et z un élément au milieu de X n'appartenant pas à X. X étant connexe pour la dépendance il existe nécessairement une dépendance $x \rightarrow y$ entre deux éléments de X qui relie les parties à gauche et à droite de z. Il existe une chaîne de dépendances entre le nœud à l'infini et z. Aucun des éléments de cette chaîne ne peut appartenir à X et l'une des dépendances de cette chaîne coupe $x \rightarrow y$, d'où P3. Enfin, si on a P3, nécessairement l'une des deux dépendances qui se coupent est non projective et on à P1. Nous avons montré l'équivalence des trois propriétés.

11.18 Flux de dépendances



Le flux de dépendances en un point de l'ordre linéaire est l'ensemble des dépendances qui relient un nœud à gauche à un nœud à droite de ce point. Le flux de dépendance est défini pour un arbre linéairement ordonné. Ce qu'on appelle un point de l'ordre linéaire est une position sur la ligne que forment les mots lorsqu'on les aligne de gauche à droite dans l'ordre linéaire. On s'intéresse généralement au flux entre deux mots successifs, c'est-à-dire au flux inter-mot, mais on peut aussi regarder le flux au-dessus d'un syntaxème donné.

(35) Flux inter-mot

Dans notre figure, le flux en un point de la phrase est l'ensemble des dépendances qui coupent le trait vertical en ce point. Le flux entre deux phrases est vide (à moins que l'on considère aussi des dépendances entre phrases). En toute

position inter-mot de la phrase, le flux contient au moins une dépendance. Le flux entre *de* et *gens* ou entre *passer* et *Noël* contient deux dépendances.

La première propriété du flux est qu'il semble être naturellement **borné**, c'està-dire que le nombre de dépendances qui appartiennent simultanément au flux en n'importe quel point de la chaîne parlée ne dépasse jamais une certaine valeur. De ce point de vue, il faut distinguer deux types de configurations : les dépendances en bouquet, qui partagent une extrémité commune, et les dépendances disjointes. Les dépendances disjointes correspondent à des enchâssements centrés (angl. *center embeddings*) de syntagmes, qui s'avère beaucoup plus coûteux pour le traitement que la présence de plusieurs dépendants sur un même élément.

(36) Dépendances en bouquet Dépendances disjointes

Les études faites sur des corpus annotés dans plus de 70 langues montrent que le flux n'a jamais plus de 6 dépendances disjointes et que ce nombre n'est quasiment jamais atteint. Plusieurs expériences en psychologie ont montré que le nombre de paramètres que l'on peut conserver simultanément dans sa mémoire immédiate dépasse rarement 7 (voir le fameux article du psychologue George Miller de 1956 intitulé The magical number seven, plus or minus two : Some limits on our capacity for processing information). Par exemple, si vous montrez pendant une fraction de seconde à une personne un écran noir avec 5 ou 6 points blancs disposés aléatoirement, elle devrait être capable de vous dire précisément le nombre de points qui sont apparus à l'écran. Mais si vous faites la même expérience avec 8 ou 9 points, cela devient beaucoup plus difficile. On peut penser que la même contrainte agit sur le flux de dépendances et rend compte du fait que nous ne sommes pas capable de traiter simultanément plus de 6 dépendances disjointes lorsque nous écoutons une phrase et l'analysons au fur et à mesure de son écoute. La même contrainte agit lorsque nous produisons un énoncé et que nous devons gérer le flux de dépendances pour assurer la cohésion de notre propos.

Nous allons maintenant préciser les liens entre le flux de dépendances et l'analyse d'un énoncé. Pour un arbre projectif, le flux est naturellement ordonné. Deux dépendances sont dites concomitantes si elles appartiennent ensemble au flux à un moment donné. Lorsque l'arbre est projectif, deux dépendances concomitantes vérifient toujours la propriété suivante : l'un des dépendances couvre entièrement l'autre. On peut donc **ordonner le flux** de la plus petite dépendance à la plus grande et voir l'ensemble des dépendances à traiter à un moment donné comme une PILE DE DÉPENDANCES (ici pile fait référence à une pile d'assiettes). Cette idée a été exploitée pour le calcul automatique d'un arbre dépendance. Nous considérons que l'analyse d'une phrase se fait **incrémentalement**, c'est-

à-dire en traitant les mots les uns après les autres dans leur ordre linéaire. À chaque mot traité, de nouvelles connexions sont effectuées et le flux de dépendances est mis à jour. Autrement dit, traiter un nouveau mot consiste à chercher ses connexions à gauche (c'est-à-dire parmi les mots déjà traités) et à introduire dans le flux ses connexions potentielles à droite. Dans l'absolu, le nombre d'information à traiter devrait grossir au fur et à mesure que de nouveaux mots sont considérés. C'est là qu'intervient la projectivité : en raison des contraintes de projectivité, les connexions potentielles qui figurent dans le flux sont ordonnées. Ce sont les dernières connexions entrées dans le flux qui doivent être traitées les premières, d'où l'idée de traiter l'ensemble des dépendances comme une pile d'assiettes : la dernière dépendance potentielle entrée dans le flux est posée sur le haut de la pile. Lorsqu'un nouveau mot est traité, seul le haut de la pile est considéré et le mot courant peut ou non accepter la connexion potentielle qui s'y trouve. S'il veut accéder à une autre connexion potentielle, il faut que la connexion potentielle qui est sur le dessus de la pile soit supprimable et supprimée. La procédure d'analyse que nous venons de décrire est appelée l'ANALYSE EN FLUX.

On peut encore noter que les positions inter-mot et les dépendances se correspondent une à une. En effet, un arbre de dépendance à n nœuds contient n-1 dépendances et, s'il est linéairement ordonné, il possède n-1 positions inter-mot. En tout point de l'ordre linéaire où le fux est projectif, on peut identifier une plus petite dépendance couvrant ce point que l'on appelle la Voûte et associer ainsi cette position inter-mot à cette dépendance. Il est intéressant de noter que cette dépendance donne une information importante sur la position inter-mot en tant que frontière. Reprenons l'exemple au début de cet encadré. La voûte de la position inter-mot entre gens et aimeraient est la relation sujet entre aimeraient et beaucoup; la position est donc la frontière droite du sujet.

11.19 Linéarisation projective

Nous allons maintenant étudier l'ordre des mots, c'est-à-dire la correspondance entre la structure de dépendance syntaxique et l'ordre linéaire, en nous plaçant dans la situation où nous voudrions ordonner un arbre de dépendance d'une langue donnée. Nous commençons par le cas, plus simple, où la linéarisation de l'arbre de dépendance donne un arbre ordonné projectif.

La projectivité contraint fortement les linéarisations possibles d'un arbre de dépendance. Elle contraint chaque nœud de l'arbre de dépendance syntaxique à rester proche de son gouverneur et donc à se placer par rapport à son gouver-

neur et à ses co-dépendants (c'est-à-dire les éléments qui dépendent du même nœud que lui).

Nous allons commencer par montrer comment le placement par rapport au gouverneur combiné avec la projectivité permet de spécifier un certain nombre d'ordres linéaires possibles, puis nous étudierons le placement par rapport aux co-dépendants dans les sections suivantes. Nous terminerons ce chapitre en étudiant la linéarisation non projective, pour laquelle le placement ne se fait plus par rapport au gouverneur.

Reprenons notre exemple favori:

(37) Beaucoup de gens aimeraient passer Noël en Laponie.

Indiquons pour chaque type de mot comment il se place par rapport à son gouverneur en français :

- le sujet précède le verbe dont il dépend;
- l'objet suit le verbe dont il dépend;
- les compléments propositionnels se placent après leur gouverneur;
- le complément d'une préposition se place après celle-ci.

De telles règles sont appelées des règles de précédence linéaire. Elles peuvent être énoncées en termes de dépendances ou de constituants, peu importe la granularité de la combinaison que nous considérons pour un type de connexion donnée.

En projetant ces règles sur l'arbre de dépendance de la phrase, on obtient la structure suivante où est indiqué, pour chaque dépendance, si le dépendant se place avant (<) ou après (>) le gouverneur :

Figure 11.8

Pour ordonner linéairement les nœuds de l'arbre, on peut par exemple parcourir l'arbre à partir de sa racine en plaçant les nœuds au fur et à mesure. On place donc d'abord le verbe *aimeraient*, puis ses deux dépendants, l'un à gauche (*beaucoup*) et l'autre à droite (*passer*) :

(38) beaucoup aimeraient passer

On place ensuite le dépendant de *beaucoup* à sa droite. La projectivité bloque une linéarisation telle **beaucoup aimeraient de gens passer*, qui respecte pourtant les règles de placement dépendant-gouverneur. En effet, *de gens* doit se passer après *beaucoup*, mais il ne peut pas aller au-delà de *aimeraient*, car *aimeraient* n'est pas dans la projection de *beaucoup*. On a donc nécessairement :

(39) beaucoup de gens aimeraient passer

On place ensuite les dépendants de *passer* à sa droite. À défaut de règles ordonnant les co-dépendants, on obtient deux ordres possibles : notre phrase de départ et la phrase *Beaucoup de gens aimeraient passer en Laponie Noël* (sur laquelle nous reviendrons dans l'encadré ?? sur les *Préférences dans l'ordre*).

Comme on le voit, la projectivité et des règles indiquant l'ordre dépendantgouverneur permettent de contrôler de manière assez précise l'ordre des mots. Nous allons voir comment compléter ces règles de base.

11.20 Linéarisation des co-dépendants

Dès qu'un nœud a plus de deux dépendants, l'un des dépendants ne pourra pas être accolé à son gouverneur. Considérons l'exemple suivant :

(40) Le petit garçon ne lui prêtera pas son autre gros ballon.

Voici son arbre de dépendance (où nous indiquons les seules fonctions pertinentes pour la suite de discussion, à savoir *sujet* et *objet*):

(Remarque : nous décidons dans cet arbre de prendre le nom comme tête du groupe substantival. Comme nous l'avons expliqué dans le Chapitre 9, le déterminant comme le nom ont des propriétés de tête. Le choix du déterminant comme tête poserait encore moins de problème pour la linéarisation.)

L'ordre dépendant-gouverneur n'est pas suffisant pour contraindre correctement les ordres possibles : il est également nécessaire d'ordonner entre eux les co-dépendants d'un même nœud qui se placent du même côté de leur gouverneur commun.

On peut envisager deux façons d'exprimer les règles d'ordre entre co-dépendants. Une première méthode consiste à considérer des règles d'ordre des co-dépendants deux à deux (ce serait à nouveau des règles de précédence linéaire). Par exemple, pour obtenir l'ordre le petit garçon et bannir *petit le garçon, on indiquera que le déterminant se place devant les adjectifs qui dépendent du même nom. Si l'on regarde maintenant, l'ordre des dépendants à gauche du verbe prêtera, on indiquera que le sujet se place avant les clitiques et que le clitique ne précède toujours les clitiques datifs (comme lui). Si on regarde plus précisément l'ordre des clitiques préverbaux du français, on voit qu'on a toujours je < ne < me < le < lui < y < en < verbe : il me le donne, je le lui donne, il n'y en a pas, etc. Comme l'ordre entre les clitiques est total, il est plus économique de donner directement l'ordre complet comme nous l'avons fait, que de donner les ordres des co-dépendants deux

à deux (étant donné qu'il y a 8 éléments à ordonner, cela nous fait 7×8/2, soit 28 couples à ordonner).

La deuxième méthode pour ordonner les co-dépendants consiste donc à donner d'un coup l'ensemble des règles d'ordre concernant un nœud et ses dépendants. Cela peut être fait en considérant que chaque nœud de l'arbre de dépendance ouvre une **boîte ordonnée** avec des positions prévues pour les différents dépendants selon leur nature et leur fonction. Nous appellerons une telle boîte un GABARIT TOPOLOGIQUE et chacune des cases de la boîte un CHAMP TOPOLOGIQUE.

Voici le gabarit qu'ouvre un verbe à l'indicatif en français :

Table 11.1

ch-pré-noyau ch-initial ch-cl-sujet ch-cl-ne ch-cl-se ch-cl-le ch-cl-lui ch-cl-y

Ce gabarit contient 7 champs préverbaux pour des clitiques (les champs ch-cl-X), un champ initial dont nous avons parlé au début de ce chapitre et qui accueille en général le sujet, un champ final qui accueille les compléments du verbe, un champ pour un verbe subordonné (notamment pour le participe des formes verbales complexes comme *a dormi*) précédé d'un champ adverbial (pour la négation *pas* entre autres) et des champs pré-noyau et post-noyau qui accueillent les éléments détachés à gauche et à droite (voir partie 6 sur la macrosyntaxe). Les éléments qui viennent se placer dans le gabarit ouvert par le verbe vont former avec lui un constituant, dit constituant topologique, que nous appelons le DOMAINE VERBAL.

Les champs possèdent des CONTRAINTES D'OCCUPATION diverses. Certains champs ne peuvent recevoir qu'un seul élément, comme les champs clitiques ou le champ initial. D'autres peuvent recevoir plusieurs éléments comme le champ final où les champs pré- et post-noyau. Certains champs sont liés comme le champ initial et le champ clitique sujet qui ne peuvent être remplis tous les deux. (Voir aussi l'encadré ?? qui suit).

La méthode avec gabarit a l'avantage que les règles d'ordre reste finalement essentiellement des règles d'ordre d'un nœud par rapport à son gouverneur. Mais au lieu d'indiquer simplement si le dépendant va à gauche ou à droite de son dépendant (< ou >), une règle d'ordre indique dans quel champ du gabarit ouvert par son gouverneur va chaque dépendant. Par exemple :



Q

- Le sujet non clitique peut aller dans le champ initial; les sujets clitiques ne vont pas dans le champ initial, car à la différence des sujets non-clitiques, ils ne peuvent pas être séparés du verbe par un insertion : Félix, je pense, prêtera son ballon vs. *Il, je pense, prêtera son ballon.
- Un pronom clitique de 1^{ère} ou 2^{ème} personne va dans le champ clitique chcl-se (il ne vous le donnera pas).
- Un complément non clitique peut aller dans le champ final. Et ainsi de suite.

11.21 Un cas d'exclusion étrange

Nous avons, dans le début de ce chapitre, considéré que l'exclusion mutuelle pouvait découler soit de l'occupation d'une même position syntaxique (*Qui Pierre la regarde-t-il?), soit d'une même position topologique (*Leider Peter popelt schon wieder in seiner Nase, voir l'encadré ?? sur Le cas des langues V2). Il existe néanmoins des cas d'exclusion qui ne relèvent ni d'une situation, ni de l'autre. Tel est le cas des clitiques te et lui (et des clitiques qui appartiennent aux mêmes paradigmes). Voici les données :

- Il te/la vendra au roi. [ici te ou la sont objets direct]
- (42)Il me/lui vendra sa fille. [ici *me* ou *lui* sont objet indirect]
- (43)Il me la vendra.
- (44)Il la lui vendra.
- (45)*Il me te vendra (vs. *Il te vendra à moi*.)
- (46)*Il te lui vendra. (vs. *Il te vendra à lui*.)

L'inacceptabilité des deux derniers exemples ne vient pas d'une exclusion syntaxique, puisque te et me/lui n'occupe pas la même position syntaxique (l'un est objet direct et l'autre objet indirect). Or il ne semble pas s'agir non plus d'une exclusion topologique, puisque me objet indirect et lui occupent des positions topologiques différentes (l'un se place avant *la* et l'autre après).

Tout se passe comme si te ou me (en tant qu'objet direct) bloquait les trois positions topologiques à la fois (celles de *te/me* objet indirect, *le* et *lui*).

il	ne	me (indirect)	le	lui	у	en	verbe

Le fait que cette exclusion ne puisse être clairement imputée ni à des règles syntaxiques, ni à des règles topologiques est un des arguments avancés pour considérer que la grammaire des clitiques préverbaux du français relèvent aujourd'hui de la conjugaison du verbe, laquelle, on le sait, est beaucoup plus irrégulière que la microsyntaxe. Nous considérons pour notre part qu'il s'agit effectivement de nanosyntaxe, sans considérer pour autant qu'il s'agit de flexion (voir Chapitre ??).

11.22 Préférences dans l'ordre

La méthode des gabarits permet de rendre compte des contraintes d'ordre rigides, comme celles des clitiques préverbaux : ici chaque élément occupe une position précise et toute variation est agrammaticale. Mais il arrive plus couramment que l'ordre entre deux co-dépendants soit libre. Tel est le cas des compléments du verbe en français. Comme le montre les exemples suivants, l'objet direct et l'objet indirect peuvent se placer assez librement l'un par rapport à l'autre :

- (47) Félix a prêté un livre à sa sœur.
- (48) Félix a prêté à sa sœur un livre qu'il aime bien.

Nous modélisons cela en indiquant que les deux compléments vont dans le même champ (que nous avons appelé le champ final). Il y a néanmoins des préférences entre les deux ordres dont la méthode des gabarits ne rend pas compte. Ces préférences dépendent de différents paramètres, les deux principaux étant le poids et la saillance communicative.

Le poids d'un constituant correspond en fait à sa taille du point de vue phonologique, c'est-à-dire à son nombre de syllabes. Un constituant qui possède un grand nombre de syllabes est dit LOURD, alors qu'un petit constituant, généralement monosyllabique, est dit léger. Plus un constituant est lourd, plus il a tendance à être éloigné de son gouverneur. Par exemple, en linéarisant l'arbre de dépendance de notre exemple favori, nous avons obtenu la phrase Beaucoup de gens aimeraient passer en Laponie Noël, qui apparaît moins naturelle que la phrase d'origine. On l'améliore nettement en alourdissant le poids de l'objet direct :

(49) Beaucoup de gens aimeraient passer en Laponie les fêtes de fin d'année.

Il est possible que joue ici une contrainte prosodique. Comme nous l'avons déjà vu dans la section 8.11 sur l'*Unité syntaxique autonomisable*, les constituants prosodiques sont des unités syntaxiques, c'est-à-dire des unités connexes pour la dépendance. Considérons l'exemple suivant, *là* est séparé de son gouverneur *passer* par un co-dépendant, *les fêtes de fin d'année* :

(50) Beaucoup de gens aimeraient passer les fêtes de fin d'année là.

Un constituant qui est séparé de son gouverneur par un co-dépendant ne peut former d'unité syntaxique avec son voisin sans inclure leur gouverneur commun. Si un constituant léger n'est pas à côté de son gouverneur, il est donc contraint soit à former une petite unité prosodique à lui-seul, soit au contraire à former une grande unité prosodique avec ses voisins qui devra inclure leur gouverneur commun *passer*. Dans les deux cas, cela nuit à l'équilibre prosodique de l'énoncé. On préférera généralement l'ordre inverse (*passer là les fêtes de fin d'année*) beaucoup moins contraignant.

La SAILLANCE COMMUNICATIVE d'un constituant est l'importance que joue ce constituant du point de vue de l'information à communiquer. Par exemple, si on parle de la Laponie et que c'est bien l'information *Noël* que l'on souhaite mettre en avant, il devient plus naturel de produire la phrase *Beaucoup de gens aimeraient passer en Laponie Noël*. Il y a par ailleurs une corrélation entre saillance et poids, puisqu'un des moyens de rendre saillante une information est de lui donner du poids.

Ce phénomène de REJET DES CONSTITUANTS LOURDS (appelé par les générativistes *heavy shift*) a été remarqué il y a bien longtemps. Ainsi, Claude Buffier, dans sa *Grammaire françoise sur un plan nouveau* de 1709 disait (p. 313) :

« Les régimes [= compléments] doivent être le plus près qu'il se peut du mot régissant; ce qui feroit pas si l'on mettoit d'abord le plus long qui éloigneroit trop le plus court. »

Nicolas Beauzée développait cette remarque dans l'entrée « Régime » de l'*Encyclopédie* en 1765 (celle-là même où il définit le complément) :

« De plusieurs *compléments* qui tombent sur le même mot, il faut mettre le plus court le premier après le mot *complété*; ensuite le plus court de ceux qui restent, et ainsi de suite jusqu'au plus long de tous qui doit être le dernier. [...] il importe à la netteté de l'expression, *cujus summa laus perspicuitas*, de n'éloigner d'un mot, que le moins qu'il est possible, ce qui lui sert de *complément*. Cependant quand plusieurs *compléments* concourent à la détermination d'un même terme, ils ne peuvent pas tous le suivre immédiatement;

et il ne reste plus qu'à en rapprocher le plus qu'il est possible celui qu'on est forcé d'en tenir éloigné : c'est ce que l'on fait en mettant d'abord le premier celui qui a le plus de brièveté, et réservant pour la fin celui qui a le plus d'étendue. »

Mais c'est à Henri Weil (**Weil1844**, p. 97-102), critiquant Beauzée (critique qui s'appliquerait à bien des analyses depuis) qu'on doit, à notre avis, l'analyse la plus pertinente :

« De plusieurs compléments qui tombent sur le même mot, donnez la forme la plus concise à celui qui suit immédiatement le mot complété et, à mesure que vous avancez, donnez aux compléments une expression plus développée et plus étendue. »

En d'autres termes, ce n'est pas parce qu'un complément est lourd qu'il doit être placé loin de son gouverneur, mais c'est parce qu'il est loin de son gouverneur qu'il doit être lourd. Weil conclut par ces mots :

« La parole est au service de la pensée, et non pas la pensée au service de la parole. »

11.23 Contraintes de poids **Q**

Une façon de rendre contre des préférences dans l'ordre des co-dépendants est de voir certaines dépendances comme pourvu d'élasticité. On peut ainsi distinguer des dépendances rigides comme la dépendance entre un clitique préverbal et le verbe dont la « longueur » est fixe et des dépendances beaucoup plus souples comme celle entre un complément non clitique et le verbe. On peut alors se représenter les contraintes de poids des constituants lourds comme véritablement exercées par les forces gravitationnelles : les différents dépendants sont comme suspendus à leur gouverneur et plus ceux-ci sont lourds, plus ils allongent la dépendance qui les relie à leur gouverneur et s'en trouvent éloignés. Considérons la phrase :

(51) Le professeur propose aujourd'hui à ses élèves un exercice sur l'ordre des mots.

Voici comment s'étirent, en fonction de leur poids, les différentes dépendance entre le verbe *propose* et ses compléments :

(52)

La position finale d'un dépendant se trouve alors déterminée par la « longueur » initiale de la dépendance (déterminée par la fonction et la nature du dépendant comme dans le modèle avec gabarit), l'élasticité de cette dépendance et différents paramètres comme le poids ou la saillance.

11.24 Les langues dites à ordre libre



Une langue est dite à ORDRE LIBRE quand une tête et ses dépendants (et notamment le verbe, son sujet et son objet) peuvent être dans n'importe quel ordre. Lorsque, au contraire, un seul ordre est possible, on dit que la langue est à ORDRE RIGIDE. Il existe entre les deux toute une gamme de langues à ordre plus ou moins libre ou plus ou moins rigide.

L'ordre des mots en français est particulièrement rigide, même s'il l'est un peu moins qu'en anglais. L'allemand a un ordre plus libre que le français. Par exemple, on peut traduire Marie poursuit Pierre par die Marie verfolgt den Pierre ou bien par den Pierre verfolgt die Marie. Les deux phrases ont le même contenu informationnel (voir chapitres

sg?

2), c'est-à-dire qu'elles décrivent le même état du monde : chaque fois c'est Marie qui poursuit Pierre et pas l'inverse. Le repérage des dépendants du verbe est effectué en allemand grâce à un SYSTÈME CASUEL, c'est-à-dire des marques sur certains mots du syntagme substantival qui n'ont pas d'autre rôle que justement d'indiquer la relation qu'entretient le syntagme avec le verbe. (Voir la notion de cas au Chapitre ??.) Ce sont avant tout les déterminants qui portent l'information casuelle en allemand. Dans notre exemple, le fait que den Pierre soit l'objet est marqué par le cas accusatif du déterminant masculin singulier den (qui s'oppose au déterminant au nominatif der).

L'ordre des mots en allemand n'est cependant pas libre au sens où seuls deux des six ordres possibles entre les trois mots correspondant à Marie, Pierre et poursuit sont possibles pour une phrase déclarative. Il existe d'autres langues, comme les langues slaves ou le grec moderne, où les six ordres des trois mots sont acceptables et désignent le même état du monde. Voici un exemple du russe :

a. Vanya prigotovil malinu (53)

'Vanya a préparé les framboises'

'Ce que Vanya a préparé, c'est les framboises'

- b. Prigotovil Vanya malinu
 'Ce qu'a préparé Vanya, c'est les framboises'
- c. Malinu prigotovil Vanya'C'est les framboises qu'a préparé Vanya'
- d. Malinu Vanya prigotovil'C'est les framboises que Vanya a préparé'
- e. Prigotovil malinu Vanya
 'Ce qu'il a préparé, c'est les framboises, Vanya'
- f. Vanya malinu prigotovil 'Vanya, c'est les framboises qu'il a préparé'

Nous avons donné ici les traductions en français lorsque ces phrases répondent à une question sous-jacente telle *Qu'a préparé Vanya*?, c'est-à-dire que *malinu* 'les framboises' est le кнèме (ce qu'on dit), tandis que *Vanya prigotovil* 'Vanya a préparé' est le тнèме (ce dont on parle).

Si on regarde les traductions françaises, on remarque que les variations entre ces phrases sont constituées de constructions qui servent à mettre en avant un élément de la phrase. La construction avec « *C'est ... que* », appelée clivée (voir le test de clivage en 3.4.9) est utilisée dans les traductions des phrases 3, 4 et 6 pour mettre en valeur un élément rhématique non verbal (ici *les framboises*). Très similaire est la pseudo-clivée « Ce que ..., c'est ... » utilisée pour les traductions des phrases 1, 2 et 5. À ces constructions s'opposent la dislocation gauche, qui sert à distinguer un élément thématique, comme *Vanya* dans la traduction de la phrase 6. Similairement, on peut aussi disloquer un élément thématique à droite comme *Vanya* dans la traduction de phrase 5. Les six ordres en russe s'accompagnent évidemment de légères différences de sens, relevant de la structure communicative (voir l'encadré suivant), similaires à celles que montrent leurs traductions en français. Il n'en reste pas moins que les six ordres sont possibles.

Notez ensuite que, abstraction faite des mots fonctionnels des différentes constructions, dans les deux langues considérées, le russe et le français, les trois mots lexicaux se trouvent toujours dans le même ordre. À cela s'ajoute que les principaux prosodèmes, qui constituent l'intonation de la phrase se ressemblent pour chacun des ordres possibles. La phrase 2 par exemple, a une courbe intonative que l'on peut schématiser comme suit, avec une intonation montante sur le thème (prigotovil Vanya) et une intonation descendante sur le rhème (malinu).

(54) Prigotovil Vanya malinu 'Ce que prépare Vanya, c'est les framboises'

Il paraît évident qu'il existe une relation entre l'existence d'un système casuel et la liberté dans le placement des arguments verbaux : il serait en effet redondant d'obliger un objet à la fois à porter un marqueur casuel d'accusatif et à occuper une place fixe. L'inverse, par contre, n'est pas tout à fait vrai. Une langue sans système casuel n'est pas forcément très restrictive quant à l'ordre des arguments verbaux. Un bel exemple est le bahasa, la principale langue parlée en Malaisie et en Indonésie, aussi appelée le malais ou l'indonésien selon ses variantes dialectales. Cette langue ne connait pas de flexion casuelle et pourtant elle permet une plus grande liberté de mots que d'autres langues sans flexion. Si l'ordre dominant est l'ordre SVO comme dans la phrase a, il est possible de postposer le sujet comme en b et d'antéposer l'objet comme en c.

- (55)a. Ali baca buku itu. Ali lire livre ce 'Ali lit ce livre'
 - b. baca buku itu Ali. lire livre ce Ali 'Il lit ce livre, Ali'
 - c. buku itu Ali baca. livre ce Ali lire 'Ce livre. Ali le lit'

On peut comparer cette situation avec celle que propose le français à l'oral, où, avec les dislocations gauches et droites, les six ordres sont possibles sans qu'il n'y ait aucune marque casuelle :

- 1. Ali, ce livre, il le lit.
- 2. Ali, il le lit, ce livre.
- 3. Il le lit, Ali, ce livre.
- 4. Ce livre, Ali, il le lit.
- 5. Ce livre, il le lit, Ali.
- 6. Il le lit, ce livre, Ali.

11.25 Ordre communicativement dirigé 🐿

Les langues à ordre rigide (voir l'encadré qui précède) sont généralement des langues où la position topologique est fortement déterminée par la fonction syntaxique et la catégorie syntaxique. Ainsi, en français, un objet direct doit nécessairement être après le verbe, à moins que ce soit un clitique, un pronom relatif

ou un pronom interrogatif. Autrement dit, en français, l'ordre linéaire est utilisé pour marqué la fonction syntaxique et permettre le repérage de l'objet par rapport au sujet. Nous dirons que de telles langues ont un ORDRE SYNTAXIQUEMENT DIRIGÉ.

A l'inverse dans **les langues à ordre libre**, la fonction syntaxique joue peu de rôle, puisque, quelle que soit la fonction syntaxique, tous les ordres sont possibles. Il ne faut pas croire pour autant que l'ordre dans ces langues soit non motivé et que tous les ordres se valent. Il existe un **principe d'économie** qui veut que les langues cherchent à en dire le maximum avec le minimum de moyens : il n'est donc même pas imaginable qu'un moyen aussi disponible que l'ordre linéaire ne soit pas exploitée par la grammaire. À quelle fin est alors utilisé l'ordre linéaire quand il ne sert pas à exprimer la fonction syntaxique?

Il semble que dès que l'ordre n'est pas syntaxiquement dirigé, il serve principalement à exprimer la STRUCTURE COMMUNICATIVE (voir l'encadré ?? sur *Les composantes du sens*) et en particulier à indiquer de quoi on parle et ce qu'on en dit (l'opposition thème/rhème) et ce que l'on souhaite contraster (la focalisation). Les langues à ordre libre sont donc généralement des langues à ordre COMMUNICATIVEMENT DIRIGÉ.

Nous avons vu, dans l'encadré qui précède, que le russe permettait tous les ordres possibles du verbe avec son sujet et son objet. Les traductions que nous avons proposées montrent que ces variations d'ordres correspondent en français à des constructions telles que le clivage ou la dislocation. En français, le clivage sert à marquer un rhème focalisé : autrement dit, dans la phrase C'est les framboises que Vanya préfère, non seulement les framboises est rhématique (ce qu'on dit, l'information qui est communiquée, c'est les framboises), mais en plus cette information est focalisée, c'est-à-dire qu'elle est contrastée avec tous les autres éléments qui pourraient occuper cette position : c'est les framboises qu'il préfère, à l'exclusion de toutes les autres choses pertinentes dans le contexte d'énonciation. La dislocation gauche quant à elle sert à marquer un thème focalisé : autrement dit, dans la phrase Les framboises, Vanya aime ça, non seulement les framboises est thématique (ce dont on parle, ce qui est le thème de l'information que je vais communiquer, c'est les framboises), mais en plus cette information est focalisée, c'est-à-dire qu'elle est contrastée avec tous les autres éléments qui pourraient occuper cette position dans le contexte d'énonciation : les framboises, il aime ça, les autres fruits rouges, c'est moins sûr.

On peut postuler qu'en russe, l'ordre linéaire de la phrase obéit au gabarit suivant :



thème-foc rhème thème

Le verbe en fonction de sa valeur communicative va venir dans un des trois champs. Les dépendants qui on la même valeur communicative que lui se placeront dans le même champ, tandis que ceux qui ont d'autres valeurs iront dans d'autres champs. La prosodie viendra souligner cette répartition avec une intonation montante sur le thème focalisé, descendante sur le rhème et généralement plate sur le thème.

Dès le 18^e siècle, l'abbé Gabriel Girard, dans son ouvrage Les Vrais principes de la langue françoise, ou la Parole réduite en méthode paru en 1747, distinguait les langues dites analogues, dont l'ordre est supposé analogue à celui de la pensée, c'est-à-dire du français, des langues dites transpositives, comme le latin ou le grec, dont l'ordre est plus libre et « transpose » l'ordre de la pensée. C'est à Henri Weil, dans sa thèse publiée en 1844 (déjà mentionné dans l'encadré ?? sur les Langues à têtes finales et langues à têtes initiales), que l'on doit d'avoir introduit les notions de thème et rhème et d'avoir noté que trois facteurs pouvaient influencer l'ordre des mots : la structure de dépendance, la structure communicative et la prosodie. En comparant le français, l'anglais, l'allemand, le turc, le chinois, le latin et le grec ancien, il remarque que certaines langues ont un ordre davantage contrôlé par la syntaxe (les langues dites à ordre rigide) et d'autres davantage contrôlé par la structure communicative (les langues dites à ordre libre).

11.26 Les autres usages de l'ordre linéaire



Nous venons de voir, dans les deux encadrés qui précèdent, les deux principaux usages de l'ordre linéaire : le marquage de la fonction syntaxique et le marquage de la structure communicative. Il existe trois autres usages de l'ordre linéaire.

Premièrement, certaines syntaxèmes peuvent avoir des contraintes de placement très particulières. En arabe, par exemple, le déterminant défini se place devant le nom (al+?awlaad+u 'les enfants $_{NOM}$ '), alors que l'indéfini se place après (?awlaad+u+n 'des enfants_{NOM}'). En français, certains adjectifs tendent à se placer devant le verbe (un petit ballon), certains se placent obligatoirement après (un ballon rouge), tandis que, pour d'autres encore, l'ordre est relativement libre (un superbe ballon; un ballon superbe). Certains adjectifs peuvent avoir selon leur acception des placements différents : un homme grand (un homme grand parmi les hommes) vs un grand homme (un homme grand dans l'humanité), un marié jeune (une personne jeune parmi les mariés) vs un jeune marié (une personne jeune dans son mariage).

Deuxièmement, les mêmes éléments lexicaux peuvent selon l'ordre déclencher des constructions différentes. Tel est le cas en russe : le numéral se place normalement devant le nom, comme dans *sto metrov* 'cent mètres', mais on peut aussi placer le numéral après le nom. Néanmoins, il s'agit d'une autre construction, car le sens change : *metrov sto* 'approximativement cent mètres'.

Troisièmement, les différences de portée peuvent également entrainer des différences d'ordre : *Le lundi, Pierre travaille à Paris* (chaque lundi Pierre est à Paris) vs. À *Paris, Pierre travaille le lundi* (quand Pierre n'est pas à Paris on ne sait pas ce qu'il fait le lundi). On peut imputer la différence de sens des deux phrases à une différence de structure communicative : dans la première *lundi* est un thème focalisé, alors que c'est *à Paris* dans la deuxième. Néanmoins, la différence de portée entraine aussi une différence de sens informationnel, puisque les deux phrases peuvent correspondre à des situations du monde différentes.

11.27 Modèle topologique

Nous allons voir de manière un peu plus précise comment fonctionne le modèle topologique basé sur les gabarits.

Nous avons donné le gabarit topologique du verbe à l'indicatif en français à la section 11.20 sur la *Linéarisation des co-dépendants*. Rappelons-le :

Table 11.2

Domaine verbal ch-pré-noyau ch-initial ch-cl-sujet ch-cl-ne ch-cl-se ch-cl-le ch-cl-lui ch-cl-y

Le gabarit topologique du nom en français est quant à lui :

Table 11.3

Domaine nominal ch-tout ch-article ch-num ch-seul ch-autre ch-adj ch-nom ch-deN chfinal

Autrement dit, le nom possède un champ final à sa droite où les compléments peuvent se placer assez librement (avec des préférences en fonction de leur poids et de leur saillance), tandis que l'ordre à gauche du nom est assez rigide : ainsi, un syntagme comme *les deux seules autres petites tables* n'accepte pas d'autre ordre que celui qu'il a. Voir l'encadré ?? sur la *Topologie du groupe substantival en français*.

Voyons maintenant comment linéariser un arbre de dépendance. Reprenons l'exemple de la section 11.20 sur *La linéarisation des codépendants* :

Figure 11.10

Nous procédons en parcourant l'arbre de dépendance à partir de la racine. On commence donc par considérer le verbe *prêtera* qui va ouvrir un domaine verbal. Les règles d'ordre indiquent pour chaque dépendant dans quel champ il peut aller en fonction de sa nature et de sa fonction : un sujet peut aller dans le champ chinitial, un clitique dans un champ ch-cl-X, etc. On obtient au final la configuration suivante (où les champs non remplis ne sont pas représentés) :

Table 11.4

garçon	ne	lui	prêtera	pas	ballon

Chaque mot va ensuite ouvrir dans la position qu'il occupe un constituant topologique qui accueillera ses dépendants. Le gabarit du constituant dépendra de la nature de l'élément qui l'ouvre et du champ qu'il occupe (voir, pour le rôle joué par le champ, le cas des adjectifs du français dans l'encadré ??, et celui des verbes en allemand dans l'encadré ?? sur *La structure topologique de l'allemand*). Ainsi les noms *garçon* et *ballon* vont-ils ouvrir des domaines nominaux, pour accueillir leurs dépendants, et donner l'ordre final suivant :

Table 11.5

le petit **ga**rçon lui prêtera pas son autre gros ballo

On notera qu'en plus de l'ordre linéaire, nous avons obtenu une structure avec des constituants emboîtés, que nous appelons la STRUCTURE TOPOLOGIQUE de la phrase.

La structure topologique est un Arbre de constituants ordonné, c'est-àdire un arbre de constituants (voir la section 10.3 sur *Arbre de constituants majeurs*) avec en plus un ordre linéaire sur les fils de chaque nœud (voir l'encadré ?? sur *Arbre de constituants ordonné*). Dans le cas de la structure topologique, l'ordre

linéaire sur les fils d'un nœud est donné par le gabarit du constituant, qui est une liste linéairement ordonnée de champs.

Figure 11.11 - Arbre topologique

Notons pour conclure que le modèle topologique n'utilise que des RÈGLES LO-CALES, c'est-à-dire des règles qui ne prennent en compte que des portions bien délimitées des structures mises en correspondance. En effet, les règles d'ordre ne concernent qu'un gouverneur et ses dépendants.

11.28 Topologie du groupe substantival en français 🖟

Dans le gabarit pour le nom proposé à la section 11.27 figure un champ ch-deN, juste après le champ qu'occupe le nom. Il existe en effet une contrainte d'ordre qui concerne les compléments en de N (préposition de suivi d'un nom nu, c'est-à-dire sans déterminant). Prenons l'exemple des dépendants d'une chemise. Nous pouvons modifié une chemise par pour homme, en coton ou pour le sport. Chacun de ces trois compléments peut aussi être réalisé par un complément en de N: d'homme, de coton, de sport. Contrairement à leurs équivalents, les compléments en de N supporte très mal d'être séparé du nom: alors qu'on dira une chemise orange en coton, on dira difficilement ?? une chemise orange de coton (on préfère une chemise de coton orange) Il semble également difficile de réaliser deux compléments en de N (?? une chemise de coton de sport). Quand on réalise un complément en de N avec un autre complément prépositionnel, le complément en de N doit précéder l'autre: ?? une chemise en coton de sport vs une chemise de sport en coton; ?? une chemise pour le sport de coton vs une chemise de coton pour le sport.

Pour les autres dépendants postposés au nom (adjectifs, compléments prépositionnels, participiales, relatives), nous considérons qu'ils vont tous dans un même champ, car leur ordre est relativement libre et semble suivre avant tout des contraintes de poids. Ainsi un adjectif se placera-t-il plutôt avant un complément prépositionnel (*un mur ancien en pierre*), mais pour peu que cet adjectif soit un peu plus lourd, il se place sans difficulté après le complément prépositionnel (*un mur en pierre très ancien*).

L'ordre des éléments antéposés au nom est lui très contraint. De plus, seul des groupes adjectivaux léger peuvent être antéposés : un haut mur, un très haut mur, mais *un haut de trois mètres mur (voir l'encadré ?? sur les Contraintes de poids). Du point de vue du modèle topologique, cela signifie que l'adjectif haut n'ouvre pas un domaine adjectival complet lorsqu'il est antéposé, mais un



domaine réduit où ne figure pas le champ final qui accueille normalement les compléments de l'adjectif:

Domaine adjectival							
ch-adv	ch-adjectif	ch-final					
Domaine adj. réduit							
ch-adv	ch-adjectif						

un mur haut de trois mètres un très haut mur

Quant à l'ordre relatif des adjectifs antéposés, évoqué à la section 11.27, il semble aller des éléments à valeur référentielle (l'article qui indique si le référent est connu ou pas, le numéral qui donne le nombre de référents, seul et autre qui permettent de positionner ce référent dans l'ensemble des référents potentiels) à des adjectifs qualificatifs, qui précise la nature du référent.

11.29 Arbre de constituants ordonné 🕏



Contrairement au cas des arbres de dépendance ordonné dont les nœuds sont totalement ordonnés, l'ordre sur les nœuds d'un arbre de constituants est partiel, puisque deux constituants qui sont dans une relation de partie-tout ne peuvent être dans une relation de précédence l'un par rapport à l'autre et que donc une telle spécification d'ordre ne serait pas pertinente.

Il est possible en théorie de considérer des arbres de constituants avec n'importe quel ordre sur les feuilles de l'arbre (les constituants minimaux), mais nous ne sommes en pratique intéressés que par les arbres de constituants ordonnés dont tous les constituants sont continus (ce qui est le cas des arbres de constituants topologiques et de la plupart des arbres syntaxiques considérés par les grammaires de constituants). Nous restreignons donc la notion d'arbre de constituants ordonné à ces seules structures.

Grâce à la restriction que nous venons d'énoncer, les arbres de constituants ordonnés possèdent deux propriétés remarquables que nous allons étudier maintenant.

Premièrement, l'ordre linéaire sur les fils de chaque nœud interne de l'arbre induit un unique ordre linéaire sur les feuilles de l'arbre. Et réciproquement, puisque l'ordre linéaire sur les feuilles suffit à récupérer l'ordre linéaire sur les fils de chaque nœud interne. La preuve est relativement simple : pour ordonner

deux feuilles, il faut rechercher leur premier ancêtre commun, lequel est unique. L'ordre sur les deux feuilles considérées est le même que celui des deux fils de leur ancêtre commun qui se trouvent sur les deux branches qui mènent à elles. Par exemple, pour ordonner *garçon* et *ne* dans l'exemple de la section précédente, il faut remonter à leur ancêtre commun qui est le nœud du domaine verbal. On en déduit que *garçon* est avant *ne*, car le domaine nominal est avant *ne*.

Deuxièmement, on peut voir les « nœuds » de l'arbre de constituants ordonné non pas comme de simples nœuds, mais comme des segments ayant un début et une fin. N'oublions pas que les nœuds d'un arbre de constituants sont des constituants d'un énoncé et que les constituants d'un énoncé sont des segments continus de cet énoncé. On peut alors représenter les constituants comme des arcs allant d'un nœud début à un nœud fin. Lesquels nœuds correspondent en fait aux positions linéaires inter-mots (si les constituants minimaux sont les mots). Voici pour la phrase *Un chat poursuit le chien* son arbre de constituant X-barre usuel, dans la représentation traditionnelle où **l'ordre est implicite**, c'est-à-dire qu'il est représenté par la position relative des nœuds sur l'axe horizontal, mais n'est pas explicitement encodé par des relations de précédence entre nœuds :

(57) Arbre de constituants ordonné (hiérarchie explicite, ordre implicite)

Voici la représentation la plus simple où l'ordre est explicite. Ici chaque constituant est représenté par un arc qui relie son début à sa fin. Les nœuds représentent alors des positions sur l'axe temporel. La hiérarchie n'est plus représentée de manière explicite, mais peut se déduire de la position relative des arcs sur l'axe vertical. Il est possible de combiner les deux représentations.

(58) Arbre de constituants ordonné (hiérarchie implicite, ordre explicite)

11.30 Grammaire de réécriture

L'une des premières modélisations mathématiques de la grammaire est la GRAM-MAIRE DE RÉÉCRITURE proposée par Noam Chomsky dans *Structures syntaxiques* en 1957 (voir l'encadré ?? sur *Calcul symbolique et grammaires catégorielles* pour les premières grammaires formelles). Notre modèle topologique, au travers de ses gabarits, prolonge ce formalisme.

Exemple

Bien qu'on puisse utiliser les grammaires de réécriture pour définir des grammaires de dépendances, il s'agit au départ d'un formalisme introduit pour écrire





des grammaires de constituants. Voici un exemple pour ceux qui ne connaissent pas encore ce formalisme. Les règles (dites de réécriture) sont de la forme :

(59)
$$P \rightarrow GN \ GV \ V \rightarrow attrape \mid poursuit$$

 $GV \rightarrow V \ GN \ N \rightarrow chat \mid chien$
 $GN \rightarrow D \ N \ D \rightarrow le \mid un$

Avec de telles règles, on peut produire des phrases telles que un chat poursuit le chien ou le chien attrape le chat. Les règles de droite sont des Règles Lexicales (par exemple attrape ou poursuit sont des constituants de type V). Les règles de gauche sont des Règles SYNTAGMATIQUES qui disent comment un constituant se décompose. Par exemple, la règle P → GN GV dit qu'une phrase P peut se décomposer en un groupe nominal GN suivi d'un groupe verbal GV. Lue autrement, elle dit comment les constituants se combinent : la combinaison d'un GN suivi d'un GV donne un P. La règle $P \rightarrow GN$ GV ne dit pas seulement que P se décompose en GN et GV, mais aussi que GN précède GV et que l'extension de P est égale à la somme des extensions de GN et GV.

On peut réinterpréter les règles syntagmatiques comme des STRUCTURES ÉLÉ-MENTAIRES, c'est-à-dire des portions de la structure finale, un arbre de constituants ordonné dans notre cas, qui, assemblées, permettent de reconstituer la structure finale. (Nous utilisons les deux modes de représentation d'un arbre de constituants ordonné présentés dans l'encadré précédent).

La règle de réécriture P → GN GV vue comme une structure élémentaire (60)

Avec les règles proposées ici, on peut générer l'arbre de constituants de la phrase Un chat poursuit le chien donné dans l'encadré qui précède. Chaque portion élémentaire de l'arbre, composée d'un nœud et de ses fils, correspond exactement à une règle de réécriture. Ainsi le sommet de l'arbre avec le nœud P et ses deux fils GN et GV est validé par la règle $P \rightarrow GN$ GV.

Critiques

Nous souhaitons faire trois critiques concernant une grammaire de constituants comme celle qui précède (qui sert toujours de base à de nombreux modèles linguistiques).

Passons rapidement sur la première puisqu'elle a déjà fait abondamment l'objet de ce livre : nous pensons qu'il est plus judicieux de penser la syntaxe d'une langue en termes de connexions et de dépendances qu'en termes de constituants (voir le Chapitre 8 sur la connexion où nous discutons les problèmes que cela pose de vouloir privilégier une fragmentation parmi toutes celles possibles), bien que, comme nous l'avons montré (Chapitre 10), les deux approches sont en grande partie équivalentes.

La deuxième critique concerne le fait de vouloir exprimer simultanément la sous-catégorisation (le fait qu'une phrase P peut être décomposée en un GN et un GV) et l'ordre linéaire (le fait que GN < GV). Ceci peut encore se comprendre pour des langues à ordre rigide comme l'anglais, mais est difficilement justifiable pour des langues à ordre libre, où l'ordre linéaire dépend peu des relations syntaxiques (voir l'encadré ?? sur *Ordre communicativement dirigé*). Cette critique a été faite dès les années 1970 et les modèles qui ont émergé au début des années 1980, notamment GPSG (*Generalized Phrase Structure Grammar*), la grammaire syntagmatique généralisée de Gerald Gazdar, bien que toujours basée sur une grammaire de réécriture, séparaient clairement les RÈGLES dites DE DOMINANCE IMMÉDIATE des RÈGLES DE PRÉCÉDENCE LINÉAIRE entre constituants.

La troisième critique permet de saisir une différence essentielle entre les grammaires dites syntagmatiques (*Phrase Structure Grammars* ou PSG) et les grammaires topologiques. Les règles de précédence linéaire des PSG sont des règles de précédence entre constituants, comme par exemple GN < GV. Or le sujet du verbe n'est pas nécessairement un groupe nominal (*Qu'il vienne me surprendrait; Partir maintenant serait plus judicieux*). Il faut donc distinguer la catégorie et la fonction. On peut remplacer cette règle par Sujet < V. C'est déjà mieux mais les grammaires topologiques vont encore plus loin. Comme nous l'avons déjà souligné, en français, le sujet n'est pas toujours devant le verbe, mais pour qu'il n'en soit pas ainsi il faut qu'un autre élément vienne occuper sa place. Il ne faut donc pas seulement déconnecter la catégorie et la fonction, mais aussi la fonction et le champ. Une grammaire topologique décompose la règle Sujet < V en deux règles : 1) il existe un champ préverbal qui doit être rempli par un et un seul élément et 2) le sujet peut occuper le champ préverbal.

11.31 Projection topologique

Comme nous venons de le voir, le modèle topologique, en plus d'assurer la linéarisation de l'arbre de dépendance, lui associe un arbre de constituants ordonné, la structure topologique. En cas de linéarisation projective, cet arbre est en fait l'arbre de constituants syntaxiques plat obtenu à partir des projections maximales des nœuds de l'arbre de dépendance (voir la section 10.5 sur *Arbre de constituants plat*).

Nous allons maintenant nous intéresser à la linéarisation non projective. Dans ce cas, certaines projections maximales ne sont pas continues. Nous allons voir comment associer un arbre de constituants ordonné avec des constituants continus à un arbre de dépendance, qu'il soit projectif ou non.

Pour chaque nœud x d'un arbre de dépendance ordonné, nous définissons la PROJECTION TOPOLOGIQUE de x comme le plus grand **segment continu** contenant x qui corresponde à une portion connexe de l'arbre de dépendance ayant x pour racine.

Lorsque l'arbre de dépendance est projectif, la projection maximale de chaque nœud est continu et donc la projection topologique d'un nœud est sa projection maximale. Mais, dès qu'un arbre de dépendance est non projectif, la projection topologique de certains nœuds n'est plus leur projection maximale. Reprenons l'exemple de la section ??:

(61) Une belle est entrée qui voulait les acheter.

La projection topologique de *une belle* est *une belle*, alors que sa projection maximale est *une belle qui voulait les acheter*. Par contre, la projection topologique de *est entrée* est bien la phrase entière et donc la projection maximale de *est entrée*.

Les projections topologiques associées à un arbre de dépendance ordonné donné forment un arbre de constituants ordonné. Nous appelons cet arbre l'Arbre de constituants topologiques ou plus simplement l'Arbre topologique induit par l'arbre de dépendance ordonné.

Arbre de dépendance et arbre de constituants syntaxiques induit

Figure 11.12

Figure 11.13 – Arbre de dépendance ordonné et arbre de constituants topologiques induit

Lorsqu'un élément n'est pas dans la projection topologique de son gouverneur syntaxique, la dépendance qui l'unit à son gouverneur est nécessairement **non projective** (voir l'encadré ?? sur l'Équivalence des définitions de la projectivité).

Une étude plus fine de la topologie d'une langue peut amener à considérer une structure topologique différente de celle de l'arbre topologique induit. Cela vient du fait que l'arbre topologique induit est calculé à partir de la seule structure syntaxique d'une phrase isolée et pas du système que forme l'ensemble des phrases d'une langue. Néanmoins pour une phrase comme celle de notre exemple, l'arbre topologique induit correspond bien à la structure topologique.

11.32 Émancipation

La comparaison entre l'arbre topologique induit et l'arbre de constituants syntaxiques induit nous amène à introduire de nouveaux concepts. Redonnons les deux arbres pour notre exemple :

Figure 11.14

Comme on le voit, la relative *qui voulait les acheter* n'est pas à la même position dans les deux arbres : dans l'arbre syntaxique, la relative est un nœud fils du groupe substantival dont *la belle* est la tête, alors que dans l'arbre topologique, elle est un nœud fils de la phrase entière dont *est entrée* est la tête. Autrement dit, la relative ne s'est pas positionnée par rapport à son gouverneur syntaxique *la belle*, mais par rapport au gouverneur de celui-ci, *est entrée*.

Un élément qui ne se place pas par rapport à son gouverneur, mais par rapport à un ancêtre plus éloigné est dit ÉMANCIPÉ. L'ancêtre par rapport auquel se place un élément donné est appelé son нôте тороlоgique.

Si l'on prend maintenant le point de vue de l'hôte topologique, nous dirons qu'un élément qui vient se placer par rapport à lui sans être un dépendant syntaxique est un VISITEUR TOPOLOGIQUE.

11.33 Non-projectivité en français

Nous allons présenter les principaux cas de non-projectivité en français et voir comment les modéliser.

1) Montée des clitiques.

- (62) a. Zoé lui a parlé.
 - b. Zoé y est sensible.
 - c. Zoé le fait appeler par un ami.
 - d. Zoé en a envie.
 - e. Zoé en connaît la moitié.

Dans tous ces exemples, le clitique se place sur le verbe fini, alors qu'il est régi par un dépendant de ce verbe, qui peut être un participe dans une forme verbale complexe (a parlé), un adjectif dans une construction copulative (est sensible), un infinitif dans une construction causative (fait appeler), un nom prédicatif dans une construction à verbe support (a envie) ou même un objet direct dans une construction transitive (connaît la moitié).

Autrement dit, le gouverneur syntaxique du clitique n'offre pas de place au clitique, forçant son émancipation. Un participe passé, par exemple, ne peut jamais accueillir des clitiques, qu'il dépende d'un verbe, comme précédemment, ou qu'il dépende d'un nom (*le livre donné à Zoé* vs **le livre lui donné*). Il va donc ouvrir un constituant avec un gabarit différent de celui d'un verbe fini :

Table 11.6

Domaine verbal réduit ch-adv ch-verbe ch-adv ch-vb-sub ch-final

Un infinitif peut normalement accueillir des clitiques (*Zoé veut l'appeler*), mais pas quand il est dans la construction causative. Il ouvrira dans ce cas un constituant réduit comme celui d'un participe passé.

- 2) Comparatif et superlatif.
- (63) a. un plus beau livre **que le mien**;
 - b. le plus beau livre du monde.

Dans ces constructions, le complément à droite du nom dépend de *plus* comme le montre le changement complet de sens lorsque celui-ci est supprimé (**un beau livre que le mien; **le beau livre du monde), ainsi que la possibilité de le séparer du nom (ce livre est plus beau que le mien; le livre le plus beau du monde).

Figure 11.15

Le complément du superlatif va donc s'émanciper, non seulement du constituant ouvert par son gouverneur *plus*, mais aussi du constituant ouvert par le gouverneur de celui-ci, *beau*, (c'est-à-dire d'un domaine adjectival réduit, voir l'encadré ?? sur la *Topologie du groupe substantival en français*) et venir ainsi se placer dans le champ final du domaine nominal. Ceci est à contraster avec les compléments de l'adjectif, qui ne peuvent s'émanciper de la même façon (*un livre beau à pleurer* vs. **un beau livre à pleurer*).

- 3) Extraction.
- (64) L'année prochaine, je pense que Zoé habitera ailleurs.
- (65) A quel endroit Zoé a-t-elle envie d'habiter?
- (66) l'endroit où il est probable que Zoé habitera.

Dans ces exemples, un complément d'un verbe subordonné, ici habiter/habitera, est placé en tête de la proposition principale : un tel complément est dit extrait. Nous consacrerons l'entièreté du Chapitre ?? à l'étude de ce phénomène, notamment lorsqu'il met en jeu des pronoms interrogatifs ou relatifs. Le premier exemple, qui illustre la TOPICALISATION d'un complément circonstanciel, peut déjà être modélisé avec les outils introduits dans ce chapitre : il s'agit d'une émancipation du complément qui vient se placer dans le champ pré-noyau ouvert par un verbe à l'indicatif. Ce champ peut également accueillir des COMPLÉMENTS dits DISLOQUÉS (car ils sont repris par un pronom) comme Zoé dans l'exemple suivant :

- (67) **Zoé**, l'année prochaine, je pense qu'elle habitera ailleurs.
- (68) L'année prochaine, **Zoé**, je pense qu'**elle** habitera ailleurs.

On note que le champ pré-noyau peut accueillir plusieurs éléments (ici le complément disloqué *Zoé* et le complément topicalisé *l'année prochaine*) et que l'ordre entre eux est libre.

11.34 Grammaire topologique formelle 🕏

On peut donner une version plus formelle de la grammaire topologique, à l'image des grammaires de réécriture (voir section ?? sur *Grammaire de réécriture*). Une grammaire topologique assure la correspondance entre un arbre de





dépendance et un ordre linéaire et elle construit une structure de constituants en même temps qu'elle assure cette correspondance. Nous allons présenter la grammaire topologique dans le sens de la linéarisation, c'est-à-dire comme le passage d'un arbre de dépendance (non ordonné) à un ordre linéaire. C'est juste un choix de présentation, puisqu'il s'agit d'une grammaire de correspondance qui met en relation les deux structures et peut être utilisée dans le sens de la synthèse comme de l'analyse.

Un arbre de dépendance comprend des nœuds d'une certaine catégorie reliés par des dépendances portant une certaine relation. La grammaire topologique s'appuie sur les catégories et les relations utilisées pour l'arbre de dépendance et introduit deux autres types d'éléments : des constituants topologiques et des champs. Une GRAMMAIRE TOPOLOGIQUE est donc la donnée de quatre ensembles d'étiquettes (catégories, relations, constituants, champs), d'un champ initial i et de quatre types de règles :

1. **Description des constituants (topologiques)**. Un constituant est un gabarit de places linéaires, c'est-à-dire une liste de champs. Les règles de ce type indiquent donc pour chaque constituant quelle est la liste des champs qui le compose. Ces règles se rapprochent des règles de réécriture d'une PSG (voir l'encadré ?? Grammaire de réécriture) et l'on peut utiliser un notation similaire : D \rightarrow f₁ f₂ ... f_n, où D est un type de constituant (D comme domaine) et les f_i des noms de champs. Nous utilisons la lettre f pour les noms de champs (de l'angl. field, all. Feld).

La principale différence avec les règles de réécriture est qu'un constituant n'est pas réécrit en une liste de constituants, mais en une liste de champs, auxquels sont assignés des constituants par d'autres règles.

D

- 2. Description des champs. Certains champs peuvent ne recevoir qu'un élément (le champ initial du français ou de l'allemand ou les champs clitiques), tandis que d'autres peuvent recevoir un nombre quelconque d'éléments. Certains champs peuvent rester vide, tandis que d'autres ne le peuvent pas. Les règles de description de champs associent donc à un champ une des quatre valeurs suivantes :! = 'un et un seul constituant',? = 'au plus un constituant', + = 'au moins un constituant', * = 'un nombre quelconque de constituants'.
- 3. Règle de correspondance. Ce sont les règles qui assurent réellement la li-

néarisation, en indiquant, pour une dépendance donnée $x \to y$, dans quel champ f peut aller le dépendant y. La règle dépend des catégories C_1 et C_2 de x et y et de la relation syntaxique r entre x et y (c'est-à-dire l'étiquette de la dépendance $x \to y$). Par défaut, le nœud y se place par rapport à son gouverneur x et le champ f est donc l'un des champs du constituant ouvert par x.

(69)

Une telle règle peut aussi indiquer la possibilité pour y de s'émanciper. On peut contrôler l'émancipation de différentes façons : celle que nous retenons est d'indiquer quelles frontières de constituants peut « traverser » y pour se placer dans un champ d'un constituant ouvert par un ancêtre de x.

4. Création de constituant. À chaque fois qu'un nœud x de l'arbre de dépendance est placé dans un champ (par une des règles précédentes), x ouvre à son tour un constituant pour placer ses dépendants. Une règle de ce type indique donc le type de constituant D créé en fonction de la catégorie C de x et du champ f dans lequel x est placé.

(70)

Signalons le cas particulier de la **règle d'initialisation** qui indique le type de constituant D créé en fonction de la catégorie C de la racine de l'arbre de dépendance lorsque celle-ci est placée dans le champ initial *i*.

Le processus de linéarisation est le suivant. On place la racine de l'arbre de dépendance dans le champ initial i où elle crée un constituant. Ce constituant reçoit une liste de champs qui pourront accueillir les dépendants de la racine. Ceux-ci créent à leur tour des constituants auxquels est assigné une liste de champs et ainsi de suite, jusqu'à ce qu'on arrive aux feuilles de l'arbre, qui créent un constituant dont elles seront le seul occupant. La structure topologique est bien formée si à la fin du processus aucun champ! ou + n'est vide.

La plupart des modèles linguistiques possèdent une opération similaire à l'émancipation pour traiter les constituants discontinus. La façon dont nous contrôlons l'émancipation s'apparente aux contraintes d'îlots (angl. *island constraints*) des grammaires génératives pour contrôler le mouvement. Rappelons néanmoins que le mouvement est une opération de transformation de la structure, alors que nous gérons la question dans la correspondance entre deux représentations de niveaux différents (voir encadré ?? sur *Linéarisation et mouvement*).

Dans la grammaire formelle que nous venons de présenter, seules la nature (ou catégorie) et la fonction ont été prises en compte. D'autres paramètres peuvent bien sûr être ajoutés et, en tout premier lieu, les valeurs communicatives (thème/rhème, focalisation, etc.). Dans de tels cas, la règle de correspondance prendra en compte, en plus de la catégorie et la fonction du dépendant, son rôle communicatif.

11.35 Constituants topologiques intermédiaires

Nous avons vu qu'un verbe à l'indicatif ouvre un constituant avec un grand nombre de champs. Certains éléments doivent se placer proches du verbe, alors que d'autres sont naturellement plus éloignés. On aura également noté, dans le cas du français, que les dépendants qui se placent plus près du verbe sont plus légers, notamment les clitiques, et que l'ordre à proximité du verbe est plus contraint (les champs ne peuvent accueillir plus d'un élément). Ceci nous amène à considérer plusieurs niveaux de COHÉSION entre le verbe et ses dépendants.

Un premier niveau de cohésion est le MOT : les syntaxèmes à l'intérieur du mot sont indissociables et inséparables comme nous le verrons dans la partie 4 consacrée à la nanosyntaxe. Autour du verbe s'amassent d'autres mots avec un ordre très rigide : clitiques préverbaux, adverbes comme pas, participe dépendant d'un auxiliaire. Nous appellerons cette unité, que nous étudierions au Chapitre ??, l'amas verbal. Un troisième niveau de cohésion est celui de la proposition avec les champs initiaux et finaux qui vont accueillir le sujet et les compléments du verbe. On appelle encore cette unité le DOMAINE MICROSYNTAXIQUE du verbe. Au-delà peuvent encore se placer d'autres éléments qui sont détachés prosodiquement du verbe et n'entretiennent pas nécessairement de relation de rection avec le verbe. Dans l'exemple suivant, le groupe substantif l'itinéraire que je vous indique ne porte aucune marque qui indique sa fonction et il sera nécessairement détaché prosodiquement de la proposition qui suit, bien qu'il forme indéniablement une seule assertion avec lui (voir partie 6).

(71) ah oui non là **l'itinéraire que je vous indique** hein vous en avez pour cinquante minutes (oral, service de renseignement téléphonique)

On parlera alors de DOMAINE MACROSYNTAXIQUE.

On peut considérer chacune de ces unités comme autant de CONSTITUANTS TO-POLOGIQUES INTERMÉDIAIRES. Ceci nous amène à décomposer le **domaine verbal** (voir la section 11.20 sur la *Linéarisation des codépendants*) en un enchâssement de plusieurs constituants :

Figure 11.16

Cette décomposition ne change pas grand-chose au fonctionnement du modèle topologique lui-même, si ce n'est qu'au lieu d'ouvrir un seul constituant, un verbe ouvrira successivement un domaine macrosyntaxique, un domaine microsyntaxique, puis un amas verbal. Par ailleurs dès qu'un dépendant du verbe ne sera pas placé dans le dernier constituant ouvert, il devra « s'émanciper » de celui-ci pour atteindre les constituants supérieurs.

Les avantages potentiels de ce découpage du domaine verbal sont multiples. Premièrement, en découpant ainsi le domaine verbal, on peut utiliser indépendamment ses différents morceaux. Nous pensons que dans certains cas le verbe n'ouvre qu'une partie des constituants : ainsi, comme nous le verrons au Chapitre ??, un verbe subordonné placé dans l'amas verbal de son gouverneur (dans le champ ch-vb-sub) n'ouvrira lui-même qu'un constituant de type amas verbal. Cette propriété sera également illustrée dans l'encadré ?? où nous présentons la *Structure topologique de l'allemand*. Nous verrons également au Chapitre ?? que le verbe d'une relative n'ouvre probablement pas de domaine macrosyntaxique.

Le deuxième avantage concerne le calcul de la prosodie. Les constituants intermédiaires marquent différents niveaux de cohésion et donc les frontières de ces constituants correspondent généralement à des coupures prosodiques de profondeurs différentes, la frontière du domaine microsyntaxique étant toujours clairement marquée prosodiquement (voir encadré suivant).

11.36 Topologie et prosodie **Q**

Nous avons déjà évoqué la prosodie et discuté des contraintes entre la structure syntaxique et la structure prosodique dans la section 8.11 *Unité syntaxique autonomisable*. Nous avons postulé que, à l'exception du cas des clitiques, les unités prosodiques sont des portions connexes de la structure syntaxique (c'est-à-dire des catenae pour reprendre le terme introduit dans l'encadré ?? *Les limites de la dualité*).

À ces contraintes s'ajoutent des contraintes plus topologiques qui concernent la cohésion des unités et que nous avons présentées dans la section précédente. Certains constituants topologiques comme les amas sont très cohésifs et accepte difficilement d'être rompus par une frontière prosodique. À l'inverse, les différentes unités du domaine macroyntaxique forment normalement des unités prosodiques distinctes. À l'intérieur du domaine microsyntaxique, les choses sont plus complexes et dépendent beaucoup de la structure communicative. Par ailleurs,

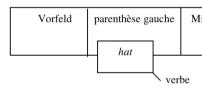
il existe des contraintes propres à chaque langue : en français, par exemple, les groupes accentuels tendent à être particulièrement équilibrés et font autour de 6 ou 7 syllabes (cette longueur est aussi dépendante du débit de parole).

En prenant en compte ces différentes contraintes, on peut faire d'assez bonnes prédictions sur les prosodies possibles. Cela dit, les locuteurs prennent pas mal de liberté et applique souvent un principe de complémentarité et de compensation qui veut que la prosodie tendra à compenser une absence de marquage syntaxique (notamment lorsque la structure est syntaxiquement ambiguë), mais pourra être plus relâchée quand la structure syntaxique est claire.

11.37 La structure topologique de l'allemand



Nous avons déjà esquissé le modèle topologique de l'allemand dans la section ?? Le cas des langues V2. La plupart des règles de l'ordre des mots en allemand se comprennent facilement en considérant qu'une phrase déclarative est consti-



tuée de 5 champs consécutifs qui forment le domaine principal.

Le verbe principal, c'est-à-dire le verbe qui porte le mode indicatif, comme l'auxiliaire hat 'a' dans notre exemple, se place en 2^e position, appelée parenthèse gauche, après l'unique constituant occupant le champ initial ou Vorfeld 'pré-champ'. La parenthèse gauche entoure, avec la parenthèse droite, le champ du milieu ou Mittelfeld. Un champ final ou Nachfeld suit la parenthèse droite. Les champs dits majeurs - Vorfeld, Mittelfeld et Nachfeld - accueillent les syntagmes nominaux ou adverbiaux. Les dépendants verbaux vont généralement dans la parenthèse droite. Si une phrase n'utilise que le domaine principal pour placer les mots, on parle d'une structure plate, par exemple :

(72)Leider hat Peter mal wieder in seiner Nase zu popeln gewagt vor nouveau dans son nez de creuser osé devant der ganzen Familie. 'Peter a malheureusement à nouveau osé se crotter toute la famille le nez devant toute la famille.'

Les parenthèses sont grisées ici. Le constituant vor der ganzen Familie 'devant toute la famille' peut aller dans le Nachfeld car il est assez lourd (et un placement dans le Mittelfeld séparerait trop les deux parenthèses). Des constituants légers comme *leider* ou *Peter* préfèrent rester dans le Vorfeld ou le Mittelfeld.

Les propositions subordonnées, complétives et relatives, ont une structure réduite, sans Vorfeld, et l'élément qui introduit la proposition, le complémenteur ou le pronom relatif, occupe la parenthèse gauche. Tous les verbes se placent dans la parenthèse droite. Les verbes recteurs se placent à droite de leur dépendant verbal.

(73) Ich glaube, [dass Peter mal wieder in seiner Nase zu popeln₃ Je crois [que Peter à nouveau dans son nez de gewagt₂ hat₁]. 'Je crois que Peter a osé se crotter le nez à nouveau.' creuser osé a]

Le placement des verbes dans la parenthèse droite peut aussi être décrit par une structure topologique : l'amas verbal. Il est composé de trois champs : le champ supérieur ou Oberfeld, le champ tête et le champ inférieur ou Unterfeld.

(74) Oberfeld champ tête Unterfeld

Sans rentrer dans tous les détails de l'amas verbal, remarquons que pour les dépendants de certains verbes, en particulier les modaux, l'Unterfeld est préférable à l'Oberfeld :

- (75) a. $^{?}$ Ich glaube, [dass Peter mal wieder in seiner Nase popeln $_{3}$ Je crois [que Peter à nouveau dans son nez creuser gekonnt $_{2}$ hat $_{1}$]. pu a]
 - b. Ich glaube, [dass Peter mal wieder in seiner Nase $hat_1 popeln_3$ Je crois [que Peter à nouveau dans son nez a creuser können $_2$]. pu]

'Je crois que Peter a pu se crotter le nez à nouveau.'

Dans l'exemple a, on observe l'ordre standard dans la parenthèse droite : l'infinitif popeln 'creuser' se place dans l'Oberfeld du participe gekonnt 'pu', lui-même placé dans l'Oberfeld de l'auxiliaire hat 'a'. On préfère en fait b, où un « ersatz » d'infinitif (all. Ersatzinfinitiv), können 'pouvoir', est réalisé au lieu du participe et où celui-ci est placé dans l'Unterfeld de l'auxiliaire, tandis que popeln 'creuser' se place à nouveau dans l'Oberfeld de son gouverneur. Cette construction est appelée l'Oberfeldumstellung 'conversion du champ supérieur'. Pour certains locuteurs de l'allemand, le verbe popeln peut se placer dans l'Oberfeld de l'auxiliaire,

non occupé par le dépendant direct de l'auxiliare. Cette nouvelle construction est appelée le Zwischenstellung 'position intermédiaire' avec l'auxilaire entre les deux verbes à l'infinitif:

(75)[c.] Ich glaube, [dass Peter mal wieder in seiner Nase popeln 3 hat 1 können 2 Je crois [que Peter à nouveau dans son nez creuser a pu] 'Je crois que Peter a pu se crotter le nez à nouveau.'

11.38 L'anglais comme langue de référence

Peu d'approches théoriques distinguent, comme nous le faisons, la structure syntaxique de la structure topologique. Nous pensons que la non-prise en compte de la topologie doit beaucoup au fait que l'anglais sert de langue de communication dans le monde scientifique et donc de langue de référence de très nombreux travaux en linguistique. Or l'anglais est, du point de vue typologique (c'est-à-dire lorsqu'on prend en compte la diversité des langues), une langue particulièrement « exotique ». En effet, l'anglais a un ordre des syntaxèmes singulièrement rigide, avec non seulement une position fixe du sujet devant le verbe, mais aussi l'obligation de placer l'objet direct avant les autres compléments. Cette particularité permet de postuler une sorte de groupe verbal (voir l'encadré ?? sur Le groupe verbal) au niveau topologique en anglais. Par ailleurs, le fait que la structure syntaxique et la structure topologique soit si fortement liée peut amener à les identifier et donc à considérer aussi un constituant syntaxique VP au niveau syntaxique, comme le font les générativistes.

Cela va plus loin encore, car les mêmes générativistes considèrent que toutes les langues devraient avoir un VP. Les langues qui autorisent le placement du sujet entre le V et l'objet, contredisant de fait l'existence d'un VP, sont alors dites non-configurationnelles. Certains linguistiques, dont Chomsky, considèrent qu'il y a bien un VP sous-jacent, mais que les syntaxèmes sont systématiquement déplacés en surface (voir l'encadré ?? sur Linéarisation et mouvement). C'est une analyse que nous rejetons totalement.

11.39 Historique de la description topologique 🖲



Nous avons déjà mentionné à deux reprises la contribution de Gabriel Girard1747 à la modélisation de l'ordre des mots. Son ouvrage contient une dizaine de règles de précédence linéaire remarquablement formalisées. Par exemple :

Q

« Première Règle. – Dans la forme expositive, le Subjectif marche ordinairement devant l'Attributif : celui-ci y précède à son tour l'Objectif et le Terminatif, lorsqu'ils sont énoncés par des expressions formelles et non simplement désignés par des pronoms personnels ou relatifs. »

Cette règle indique que, dans la phrase déclarative, le sujet (= Subjectif) se place devant le prédicat verbal (= Attributif), qui lui-même précède l'objet direct (= Objectif) et l'objet indirect (= Terminatif), à moins que ceux-ci ne soient des pronoms personnels ou relatif. Les règles suivantes indiquent que le sujet se place après le verbe dans les propositions en incise (Règle II), que le sujet peut se placer après le verbe en l'absence d'un objet direct ou quand celui-ci est un pronom (Règle III), etc.

Le modèle topologique, avec ses gabarits de places fixes, s'est développé au siècle suivant en Allemagne. Simon (Heinrich Adolf) Herling1821 propose la première théorie globale sur la structure hiérarchique des phrases complexes de l'allemand basée sur des gabarits de places fixes. Oskar Erdmann1886 poursuit le travail de Herling en énumérant le type d'éléments que ces places peuvent contenir, donnant ce que Höhle1986 propose d'appeler le système de Herling-Erdmann et qu'on appelle plus souvent aujourd'hui la théorie des champs topologiques. Le terme Feld 'champ' pour désigner ces places dans la phrase apparaît pour la première fois chez Erich Drach1937 dans un livre intitulé Grundgedanken der Deutschen Satzlehre 'Idées fondamentales de la phrase allemande' destiné aux enseignants de l'allemand comme langue maternelle et langue étrangère. Le livre de Drach prône, avec une teinte légèrement nationalisante, une émancipation de la grammaire allemande, qui était jusqu'à là sous une influence latine forte, en faveur d'une « construction d'une présentation et d'un système de règles basé sur la nature de la langue allemande ». Gunnar Bech1955 a adapté par la suite la terminologie de Drach pour décrire la structure interne du complexe verbal.

Le modèle topologique a été appliqué par Povl Skårup (??) à la description de l'ordre des mots en ancien français, montrant l'influence des langues germaniques sur l'émergence du français. Gerdes et **Kahane2006** proposent le modèle topologique pour le français présenté dans ce chapitre, qui s'appuie notamment sur les travaux sur la macrosyntaxe de Claire Blanche-Benveniste1990. Le modèle topologique a aussi été appliqué à des langues clairement non germaniques, comme dans la description par **DonohueSag1999** du warlpiri, une langue aborigène d'Australie à ordre très libre.

La question de l'ordre des mots a été peu traitée dans les premières grammaires formelles. Dans les grammaires de constituants, l'ordre linéaire est intégré à la structure syntaxique et toute variation de l'ordre des mots suppose



un changement de structure syntaxique (voir l'encadré ?? De la non-séparation des ordres au mouvement). Il faudra attendre le début des années 1980 pour que soient proposées des grammaires de constituants où les règles de précédence linéaire sont séparées des règles de sous-catégorisation, notamment avec le modèle GPSG (GazdarEtAl1985), qui donnera ensuite HSPG (PollardSag1987). Le modèle topologique de l'allemand sera formalisé pour la première fois par Andreas Kathol1995 dans le cadre de HPSG.

Dans le cadre des grammaires de dépendance, Igor Mel'čuk et Nicolaj Pertsov1987 proposent une liste très complète des constructions de l'anglais et des règles de précédence linéaire qui vont avec. Le modèle topologique sera formalisé en grammaire de dépendance simultanément par Duchier et Debusmann2001 et Gerdes et Kahane2001 (voir la formalisation proposée dans l'encadré ?? Grammaire topologique formelle).



Exercices

Exercice 1 a) Qu'est-ce que la topologie?

b) À quel endroit du modèle linguistique le modèle topologique intervientil dans un modèle stratificationnel comme la théorie Sens-Texte (voir section 3.9 *Modularité et stratification* et encadré ?? *La Théorie Sens-Texte*)?

c) Pourquoi rejetons-nous la notion d'ordre de base?

Exercice 2 Ordre préfixé et postfixé. a) On considère l'expression préfixée × 3 + × 12 5 7. Donner l'arbre de dépendance correspondant et en déduire une expression infixée et postfixée.

b) On considère l'expression postfixée 1 5 $\sqrt{+2}$ /. Sachant que $\sqrt{-1}$ (racine carrée) est un opérateur unaire (c'est-à-dire d'arité 1), donner l'arbre de dépendance correspondant à cette expression. En déduire l'écriture traditionnelle de ce nombre (qui n'est autre que le nombre d'or).

Exercice 3 Projectivité. Construire la structure de dépendance de la phrase suivante et vérifier qu'elle est non projective :

À cette heure-là, je pense qu'il est déjà parti.

Exercice 4 Théorie des graphes. Le graphe K_4 n'est pas planaire extérieur (voir encadré *?? Projectivité et planarité*). Montrer qu'il est néanmoins planaire et que donc tous les graphes de 4 nœuds ou moins sont planaires. Chercher les graphes non-planaires les plus simples à 5 et à 6 nœuds.

Exercice 5 Topologie du groupe substantival en français. On s'intéresse plus particulièrement au placement des déterminants et des numéraux en français. Comment rendre compte des données suivantes dans un modèle topologique?

- (76) les/mes/ces/deux/quelques/des/plusieurs amis sont venus
- (77) les/mes/ces deux/quelques amis sont venus
- (78) *les mes/ces/des/plusieurs amis sont venus
- (79) *amis sont venus

Exercice 6 Interrogation en français. On observe un contraste entre le fonctionnement des interrogatifs à qui et que:

- (80) À qui parle-t-elle?
- (81) À qui Marie parle-t-elle?
- (82) Que dit-elle?
- (83) *Que Marie dit-elle?

Comment expliquer cette différence de comportement dans le cadre d'un modèle topologique?



Exercice 7 Topologie de l'anglais. L'anglais possède une classe de verbes que l'on appelle les modaux qui comprend les auxiliaires BE 'être' et HAVE 'avoir' et quelques verbes à valeur modale comme CAN 'pouvoir, MUST 'devoir, WILL (auxiliaire du futur) ou WOULD (auxiliaire du conditionnel) qui ont la particularité d'être invariables, ainsi que do 'faire'. Ces verbes ont également des propriétés distributionnelles qui les distinguent des autres verbes. Premièrement, les adverbes se placent de préférence après les modaux et avant les verbes ordinaires :

- (84)Mary often calls Peter 'Marie appelle souvent Pierre'
- ?? Mary calls often Peter (85)
- Mary would often call Peter 'Marie appellerait souvent Pierre' (86)
- ?? Mary often would call Peter (87)

Deuxièmement, dans les interrogatives, le sujet se place entre le modal et le verbe, et en cas d'absence de modal, do est introduit :

- (88) Would Mary call Peter? 'Marie appellerait-elle Pierre?'
- (89)*Calls Mary Peter?
- Does Mary call Peter? 'Marie appelle-t-elle Pierre?' (90)

La position initiale peut aussi être occupée par un pronom interrogatif :

- Why would Mary call Peter? 'Pourquoi Marie appellerait-elle (91)Pierre?
- Who does Mary call? 'Qui Marie appelle-t-elle?'

Troisièmement, la négation se place obligatoirement sur un modal et donc comme dans le cas de l'interrogation, do doit être introduit en l'absence d'un autre modal:

- (93) Mary would not call Peter 'Marie n'appellerait pas Pierre'
- (94) *Mary calls not Peter
- (95) Mary does not call Peter 'Marie n'appelle pas Pierre'

 Quatrièmement, le sujet peut être inversé dans certains cas :
- (96) Never would Mary call Peter 'Jamais Marie n'appellerait Pierre'
- (97) Here are two nice people 'Ici sont deux chouettes personnes'

Rappelons également que, en anglais, l'objet direct doit précéder tous les autres compléments. De quelle façon un modèle topologique de l'anglais peut-il rendre compte de ces propriétés?

Exercice 8 Topologie de l'allemand. Les exemples que nous avons présentés dans l'encadré ?? acceptent encore d'autres ordres, comme, par exemple :

(98) In seiner Nase zu popelnhatPetermal wieder gewagt.dans son nez de creuser a Peter à nouveau

osé

'Peter a à nouveau osé se crotter le nez.'

Comment pouvez-vous intégrer cet exemple au modèle topologique de l'allemand?

Exercice 9 Ce que nous avons fait pour le français, l'allemand, le russe ou l'anglais peut être fait pour n'importe quelle langue a priori. Nous avons déjà écrit des modèles topologiques pour l'arabe, le chinois ou le wolof. Vous pouvez essayer d'écrire un fragment de modèle topologique pour la langue de votre choix et nous l'envoyer.

11.39 Historique de la description topologique



Lectures additionnelles

Comme nous l'avons déjà mentionné au Chapitre 9, on peut consulter en ligner la plupart des ouvrages anciens. On trouvera facilement **Buffier1709**, **Girard1747**, **Weil1844**, ainsi que les articles de Beauzée et Dumarsais dans l'*Encyclopédie* de Diderot et D'Alembert (voir Chapitre 9).

Les premiers travaux sur le modèle topologique sont en allemand : Erdman1886, Drach1937, Bech1955. Pour des travaux plus récents sur l'ordre des mots, nous avons mentionné GazdarEtAl1985, PollardSag1994, Mel'čuk & Pertsov1987, Kathol1995 ainsi que les articles de BresnanEtAl2007, DuchierDebusmann2001 et GerdesKahane2001; GerdesKahane2006. Sur la structure communicative et son rôle dans l'ordre des mots, nous renvoyons à Lambrecht1996 et Mel'čuk (??).

Pour les travaux typologiques sur l'ordre des mots, on consultera l'article original de **Greenberg1963** et l'étude et le travail très complet de **Dryer1992**, ainsi que les cartes et les différents articles consacrés à l'ordre des mots sur le site wals.info, une base de données typologique où sont répertoriés 192 traits pour plus de 2500 langues!

Joan Bresnan, Anna Cueni, Tatiana Nikitina, R. Harald **Baayen2007**. Predicting the dative alternation, *Cognitive foundations of interpretation*, KNAW, 69-94.

Gunnar Bech1955 Studien über das deutsche Verbum infinitum [Étude sur le verbe infinitif allemand], Band I (= Det Kongelige Danske Videnskabernes Selskab, Historisk-filosofiske Meddelelser 35, no. 2), Copenhague.

Erich **Drach1937** *Grundgedanken der deutschen Satzlehre* [Idées fondamentales de la syntaxe allemande] Diesterweg, Frankfurt am Main.

Matthew S. **Dryer1992** The Greenbergian Word Order Correlations, *Language*, 68(1), 81-138.

Denys Duchier, Ralph **Debusmann2001** Topological dependency trees: A constraint-based account of linear precedence, *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Oskar **Erdmann1886** *Grundzüge der deutschen Syntax nach ihrer geschichtlichen Entwicklung* [Caractéristiques de base de la syntaxe allemande



selon son évolution historique], Vol. 1. JG Cotta.

Gerald Gazdar, Ewan Klein, Geoffrey Pullum, Ivan Sag1985 Generalized phrase structure grammar. Harvard University Press.

Simon Heinrich Adolf **Herling1821** *Uber die Topik der deutschen Sprache*, Varrentrapp.

Kim Gerdes, Sylvain Kahane2001 Word order in German: A formal dependency grammar using a topological hierarchy, Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL).

Kim Gerdes, Sylvain Kahane2006 L'amas verbal au cœur d'une modélisation topologique du français, Lingvisticae Investigationes, 29(1), 75-89.

Joseph H. Greenberg1963 Some universals of grammar with particular reference to the order of meaningful elements, in J. H. Greenberg, Universals of language, 73-113.

Andreas Kathol1995 Linearization-based German syntax, thèse de doctorat, The Ohio State University.

Knud Lambrecht1996 Information structure and sentence form: Topic, focus, and the mental representations of discourse referents. Vol. 71. Cambridge University Press.

Igor Mel'čuk, Nicolaj Pertsov1987 Surface Syntax of English: A Formal Model within the Meaning-Text Framework, Benjamins.

Igor Mel'čuk (??) Communicative organization in natural language, Benjamins.

Carl Pollard, Ivan A. Sag1994 Head-driven phrase structure grammar, University of Chicago Press.

Povl Skårup (??) Les premières zones de la proposition en ancien français : essai de syntaxe de position. Vol. 6., Akademisk Forlag.



Corrections des exercices

Corrigé 1 a) La topologie est l'étude du placement des unités syntaxiques, c'est-à-dire l'étude de l'ordre des mots, mais aussi des syntaxèmes à l'intérieur des mots.

- b) Le modèle linguistique décrit la correspondance entre le sens et le texte. Le modèle topologique décrit la linéarisation, c'est-à-dire la correspondance entre la structure syntaxique (non ordonnée) et la chaîne linéaire. Ou dans sa version plus élaborée entre une structure de dépendance syntaxique et un arbre de constituants topologiques.
- c) Certaines langues ont un ordre dominant, comme l'ordre SVO en français. Néanmoins considérer un ordre de base revient à considérer que tous les autres ordres sont obtenus par des transformations de l'ordre base. Cela revient à dire que lorsque le sujet est après le verbe en français, celui-ci a été inversé. Même s'il peut nous arriver de conserver la terminologie et de parler de « sujet inversé », nous ne considérons pas qu'il y a eu une inversion. Nous considérons que le sujet a été placé directement dans cette position.
- **Corrigé 2** 1) Le premier symbole, \times , est la racine de l'arbre. Son premier dépendant est 3. Le symbole suivant, +, est la racine du deuxième dépendant, et ainsi de suite. La formule est donc $3 \times ((12 \times 5) + 7)$ sous forme infixée et $3 \times 12 \times 7 + \times$ sous forme postfixée.
- 2) La formule est $(1+\sqrt{5})$ / 2 sous forme infixée et / + 1 $\sqrt{5}$ 2 sous forme préfixée.
- Corrigé 3 Le syntagme à cette heure-là dépend de parti et couvre donc la racine de l'arbre pense.
- Corrigé 4 Le graphe K_4 est planaire : il suffit de prendre une des diagonales et de la faire passer par l'extérieur du carré. C'est le plus complexe



des graphes à 4 nœuds et donc tous les graphes à 4 nœuds ou moins sont planaires. Le plus petit graphe non planaire à 5 nœuds est K₅, le graphe complet à 5 nœuds. Le plus petit graphe non planaire à 6 nœuds est K_{3,3}, le graphe où 3 nœuds sont liés aux 3 autres. On vérifiera que, si on retire un seul de leurs liens, ces graphes deviennent planaires. (99)

Le mathématicien polonais Kazimierz Kuratowski a établi en 1930 la caractérisation suivante des graphes planaires : Un graphe est planaire si et seulement s'il ne peut être réduit ni à K₅, ni à K_{3 3}.

On pourra consulter les pages de la wikipédia sur les graphes planaires et planaires extérieurs.

Corrigé 5 On observe trois paradigmes de « déterminants » : 1) les, mes, ces; 2) des, plusieurs; 3) deux, quelques. Les éléments de types 1 et 3 peuvent cooccurrer, tandis que ceux de types 2 excluent les autres. Les noms communs doivent obligatoirement être accompagnés d'au moins un éléments d'un de ces trois types. Les éléments de type 1 sont les déterminants définis, ceux de type 2 les déterminants indéfinis. Quant à ceux de type 3, ce sont des quasi-déterminants, puisqu'ils peuvent se placer entre un déterminant défini et le nom et qu'ils n'ont pas de valeur intrinsèquement définie ou indéfinie, mais ils prennent une valeur indéfinie en l'absence d'un déterminant défini. On peut modéliser cette distribution en considérant deux champs topologiques : un pour les déterminant définis, suivi d'un pour les quasi-déterminants, avec la condition qu'un des deux champs au moins doit être occupé. Les déterminants indéfinis occuperaient quant à eux occuperaient les deux champs en même temps. (Ou ce qui revient au même, ils occupent le champ des quasi-déterminants, mais excluent, pour des raisons sémantiques, la cooccurrence avec un déterminant défini.)

Corrigé 6 Les pronoms interrogatifs se placent au début du domaine microsyntaxique, devant le sujet (dans Marie, à qui parle-t-elle?, Marie est détaché et l'occupe plus la position sujet). Le pronom interrogatif que est la forme faible du pronom *quoi* (comme *me* avec *moi*). C'est un clitique qui doit obligatoirement être dans l'amas verbal (comme me). Autrement dit, le pronom niterrogatif que doit à la fois être en tête du domaine microsyntaxique et dans l'amas verbal, ce qui écrase tout ce qui se trouve entre les deux et en particulier le champ initial qu'occuperait le sujet.

Corrigé 7 Les modaux de l'anglais occupent un champ topologique différent de celui des verbes. Le domaine de la phrase déclarative contient les champs suivants :

ch_initial | ch_modal | ch_adv | ch_verbal | ch_objet | ch_compl

Le sujet va dans le champ initial, sauf quand celui-ci est occupé par un pronom interrogatif, une négation ou un locatif, ou quand il reste vide dans une interrogative totale. Le champ modal peut rester vide quand le sujet occupe le champ initial et qu'il n'y a pas la négation *not* dans le champ adverbial. Le sujet inversé va dans le champ adverbial.

Corrigé 8 Cet exemple montre qu'un constituant à tête verbale, comme *in seiner Nase zu popeln* 'de se crotter le nez', peut également venir se placer dans le Vorfeld. Il n'est pas nécessaire d'aménager davantage le modèle.