

REMARQUES SUR LES BASES DE LA CONJUGAISON

Olivier Bonami

Université Paris-Sorbonne & LLF (UMR 7110)

Gilles Boyé

ERSS (UMR5610) & Université Nancy 2

La modélisation des données de la flexion met souvent en jeu une tension entre phonologie et morphologie : face à la multiplicité des réalisations d'un même lexème, on est amené, soit à postuler des représentations et des opérations phonologiques abstraites, soit à rendre plus complexe la représentation morphologique du lexème. Cet article considère un cas particulier de cette situation, la modélisation des alternances morpho-phonologiques intervenant dans la conjugaison régulière en français, qui rendent problématique la détermination d'une forme de base des lexèmes verbaux. Dans la tradition générative, ces alternances ont été mises sur le compte de formes phonologiques sous-jacentes susceptibles de réalisations transparentes ou opaques suivant le contexte flexionnel (voir entre autres Schane 1968, Plénat 1987). Nous nous basons sur des idées issues du *Minimal Generalization Learner* de Albright et Hayes (2003), et sur la notion d'espace thématique mise en place par Bonami et Boyé (2003) pour proposer une nouvelle approche de ces données : tous les lexèmes réguliers sont associés à une forme de base, mais celle-ci n'a pas le même statut morphologique pour tous les lexèmes. Spécifiquement, ce n'est pas toujours la même case de l'espace thématique qui sert de forme de base. Cette hypothèse permet de rendre compte de la conjugaison régulière sans faire d'hypothèse coûteuse sur la nature des représentations phonologiques.

1. DE LA FORME SOUS-JACENTE À L'ESPACE THÉMATIQUE

1.1. PLÉNAT (1987)

Quelle modélisation peut-on proposer pour les variations de forme que met en œuvre la conjugaison des verbes français ? Plénat (1987) tente de répondre à ces questions en poussant au bout les outils de la phonologie et de la morphologie génératives classiques : les formes de surface des passés simples et des participes passés des verbes des second et troisième groupes sont obtenues par un jeu de règles phonologiques à partir d'une forme sous-jacente unique du verbe. Ainsi le participe passé de SAVOIR /sy/ est-il obtenu à partir du radical /sav-V/ (V étant une voyelle sous-spécifiée) et du suffixe de participe /t/. Cette étude se termine sur une conclusion à la lucidité admirable :

- (1) Les thèmes de PS et de PP de l' « autre » conjugaison constituent un ensemble fini et peu susceptible de s'accroître. Dans un tel ensemble, il est fatal qu'un linguiste découvre un certain nombre de régularités : il doit au moins déceler les traces des évolutions régulières ou analogiques par lesquelles est passé le système. Mais a-t-il pour autant le droit de supposer que ces régularités sont repérées et apprises comme telles par les locuteurs, d'en faire des règles que la grammaire de chacun intériorise ?

Plénat (1987 : 138)

L'auteur conclut que seules les données psycholinguistiques sont à même de trancher la question : l'examen des erreurs des apprenants est à même de nous informer sur le fait qu'ils exploitent ou non les généralisations observables. L'exemple suivant est particulièrement frappant : parmi les 145 verbes dont l'infinitif se termine en /Cr/, où C est une occlusive, une minorité (44%) a un participe passé en /Cy/. Le caractère minoritaire de la classe, ainsi que sa petite taille, pourrait faire croire qu'il n'y a pas de

généralisation à faire ici. Pourtant, les parents de jeunes enfants constatent tous l'opiniâtreté avec laquelle les apprenants cherchent à étendre cette classe, et produisent des participes passés comme /prãdy/ pour PRENDRE ou /pẽdy/ pour PEINDRE. Quoique cette observation soit anecdotique, elle laisse penser qu'il existe une régularité inhérente à la conjugaison du français et qui veut que certains types de verbes ont, de manière régulière, un participe passé en /y/, et ce malgré la petite taille de la classe de verbes concernée.

1.2. LE MINIMAL GENERALIZATION LEARNER

La littérature psycholinguistique récente sur la flexion fourmille de propositions de modélisation de la flexion irrégulière ou semi-régulière ; la plupart des travaux s'accordent à penser que cette partie de la flexion fonctionne sur le mode associatif, les lexèmes formellement semblables tendant à se fléchir de la même manière.

Parmi les travaux de ce type ressort le modèle de *Minimal Generalization Learner* (« Apprenant de généralisations minimales » ; dorénavant MGL) de Albright et Hayes (Albright, 2002a, 2002b ; Albright et Hayes, 2002, 2003). Ce modèle est un modèle stochastique à base de règles phonologiques familières à la SPE (Chomsky et Halle 1968), ce qui rend le dialogue entre travaux linguistiques et psycholinguistiques plus aisé.

Le MGL tente de modéliser le comportement d'un sujet qui apprend à former une certaine forme d'un lexème sur la base d'une autre. L'idée générale est que face à une nouvelle paire de formes, l'apprenant construit toutes les règles phonologiques qui pourraient relier ces deux formes, puis fait un calcul de fiabilité des règles concernées en prenant en compte leur efficacité passée.

Reprenons l'exemple de l'infinitif et du participe passé en français. L'apprenant est exposé à des paires de formes comme $\langle \text{lave, lave} \rangle$ (LAVÉ), $\langle \text{finir, fini} \rangle$ (FINIR), $\langle \text{r\~{a}dr, r\~{a}dy} \rangle$ (RENDRE), $\langle \text{pr\~{a}dr, pri} \rangle$ (PRENDRE), etc. Considérons ce qui se passe quand la paire $\langle \text{r\~{a}dr, r\~{a}dy} \rangle$ est rencontrée. Toutes les règles susceptibles de relier ces deux formes doivent être considérées ; quelques exemples sont listés en (2).

$$\begin{array}{ll}
 (2) \quad \text{a. } r \rightarrow y / _ \# & \text{b. } r \rightarrow y / \left[\begin{array}{l} +\text{consonantique} \\ -\text{vocalique} \\ -\text{sonnant} \\ -\text{continu} \end{array} \right] _ \# \\
 & \text{c. } r \rightarrow y / d _ \# \quad \text{d. } r \rightarrow y / \left[\begin{array}{l} -\text{consonantique} \\ +\text{vocalique} \\ +\text{nasal} \end{array} \right] d _ \# \\
 & \text{e. } r \rightarrow y / \tilde{a}d _ \# \quad \text{f. } r \rightarrow y / \#r\tilde{a}d _ \#
 \end{array}$$

Toutes ces règles sont également utilisables pour dériver /r\~{a}dy/ de /r\~{a}dr/, mais elles ne sont pas également *précises*. À un extrême, la règle (2a) a une précision à peu près nulle, puisqu'elle donne des résultats incorrects pour l'immense majorité des verbes à infinitif en /r/, dont le participe passé n'est pas en /y/. À l'autre extrême, (2f) a une précision parfaite, puisqu'elle s'applique correctement à un verbe unique. Les cas intermédiaires ont des niveaux de précision intermédiaires.

Une fois déterminée la précision d'une règle, reste à déterminer sa fiabilité : une règle ayant une précision parfaite, comme (2f), est de peu d'utilité dans la mesure où elle n'est applicable qu'à un petit ensemble de forme. La mesure de fiabilité proposée par Albright et Hayes pénalise les règles qui ont un champ d'application étroit, même si

leur précision est grande¹. La mesure de fiabilité des différentes règles est indiquée en (3).

(3) Fiabilité des règles (2)² : $r \rightarrow y$

N°	(2a)	(2b)	(2c)	(2d)	(2e)	(2f)
Précision	9,7 %	45,5 %	51,5 %	51,8 %	72,7 %	100 %
Fiabilité	9 %	42,8 %	48,1 %	48,1 %	67,6 %	31,7 %

On voit que dans ce cas particulier, la fiabilité de la règle croît plus ou moins avec la précision du contexte, et ce malgré le fait que le nombre de cas concernés décroît simultanément. La règle la plus fiable est la règle (2e). La plus précise est (2f), mais elle est peu fiable, ne s'appliquant qu'à un seul verbe. Par ailleurs elle n'a aucun intérêt pratique, puisque la formulation de la règle est telle qu'elle ne *pourrait pas* s'appliquer à un autre verbe. Elle ne peut donc permettre à un locuteur de savoir comment fléchir un verbe nouveau.

Sur la base de ce type de calcul de fiabilité, le MGL permet de simuler le comportement d'un locuteur quand il tente de fléchir une nouvelle forme : parmi les règles applicables à cette forme que le locuteur a déjà rencontré, la plus fiable est celle qui a le plus de chances d'être appliquée. Par exemple, s'il tente de construire le participe passé de FINIR, il est très improbable que le locuteur utilise la règle (2a), dont la fiabilité est très mauvaise ; il lui préférera à coup sûr la règle $[r \rightarrow \emptyset / i_ \#]$, qui a une fiabilité de 76,6 %. En revanche, face à l'infinitif de PRENDRE /prãdr/, l'utilisation de la règle (2e), dont la fiabilité est assez bonne, est compréhensible.

¹ Le détail du calcul de fiabilité est donné dans Albright (2002, p. 40-41). Pour aller vite, la précision (*raw reliability* dans le vocabulaire d'Albright) est le rapport entre le nombre d'items correctement couverts par la règle et le nombre d'items candidats. La fiabilité (*adjusted reliability*) est obtenue en retranchant à la précision sa limite de confiance inférieure, un nombre qui augmente quand le nombre de candidats diminue ; à précision égale, une règle est donc plus fiable si elle s'applique à plus de cas.

² Les calculs reposent sur les formes verbales de BDLEX (de Calmès & Pérennou 1998), qui contient environ 6500 verbes.

Le modèle du MGL a deux traits particulièrement intéressants. D'une part, il permet de largement dépasser le débat sur la pertinence de *règles* au sens traditionnel, par opposition à une mémoire associative (souvent modélisée par un réseau de neurone) dans l'analyse de la flexion : dans un modèle stochastique comme le MGL, tous les objets manipulés sont des règles, ce qui n'empêche pas le modèle de rendre compte des effets de ressemblance de famille entre lexèmes. D'autre part, le MGL met en avant le fait qu'un même changement morpho-phonologique peut avoir des fiabilités variables en fonction des propriétés de la forme de base : comme on l'a vu plus haut le passage de /r/ à /y/ pour former le participe passé en français est plus fiable pour certaines classes de lexèmes que pour d'autres. Cette observation a permis de mettre au jour la phénoménologie des *îlots de fiabilité*, observée pour l'italien dans Albright (2002a) et pour l'anglais dans Albright & Hayes (2003) : par exemple, le fait que la règle par défaut qui veut qu'en anglais le passé d'un verbe soit formé sur la forme de base en suffixant /d/ ou /t/ n'a pas d'exceptions dans le cas où la forme de base se termine par une fricative sourde ; Albright & Hayes montrent expérimentalement que cette généralisation, qui est une sorte de cas particulier de la règle par défaut, est exploitée par les locuteurs de manière détectable.

Au delà de la portée strictement psycholinguistique de ce travail, une leçon claire pour la modélisation de la grammaire de compétence est qu'il peut exister plus d'une recette pour relier deux formes d'un paradigme. Alors que les morphologues tendent à privilégier les analyses qui associent une règle unique à la construction d'une case du paradigme (ou éventuellement, une règle par classe flexionnelle), Albright & Hayes mettent en avant la nécessité de dégager une pluralité de règles.

1.3. LES ESPACES THÉMATIQUES

Bien que le MGL constitue un changement considérable par rapport aux visions classiques du fonctionnement de la morphologie, il reste fidèle à la tradition morphophonologique générative sur un point : il est présupposé qu'il existe une unique forme phonologique qui sert de base à la construction de l'ensemble du paradigme. Alors que les analyses classiques, comme Plénat (1987), supposent que cette forme unique est une forme sous-jacente abstraite, Albright & Hayes supposent que c'est une forme de surface ; mais il reste que cette forme unique est supposée permettre la construction d'un paradigme complet³.

Pour des raisons indépendantes, nous avons proposé, dans une série de travaux récents (Bonami & Boyé 2003, *sous presse* ; Bonami, Boyé & Kerleroux à paraître), en élaborant sur Aronoff (1994), de remettre en cause l'idée de la forme de base unique en flexion. La proposition est que chaque lexème est associé à un *espace thématique*, une famille indexée de formes qui servent à construire différentes zones du paradigme.

La principale motivation de ce choix est le traitement qu'il permet de la morphologie irrégulière et en particulier de la supplétion. Quand on examine la conjugaison des verbes en français, on constate qu'il existe des zones du paradigme qui ont une grande stabilité, et ce pour tous les verbes, quelle que soit leur irrégularité par ailleurs. Comparons par exemple les formes de l'imparfait et du futur, aux première et deuxième personnes du pluriel ; quelques exemples sont donnés en (4).

(4)

³ Albright (2002b) consacre une grande partie de sa thèse à savoir comment les apprenants décident *laquelle* des cases du paradigme sert de base aux autres cases. L'idée générale est d'appliquer le MGL à chacune des hypothèses possibles et de garder le choix qui donne lieu aux règles les plus fiables.

(4)

VERBE	IMP 1PL	IMP 2PL	FUT 1PL	FUT 2PL
LAVÉ	lavjǝ	lavje	lavrǝ	lavre
CONCLURE	kǝklyjǝ	kǝklyje	kǝklyrǝ	kǝklyre
ALLER	aljǝ	alje	irǝ	ire
PEINDRE	pǝnjǝ	pǝnje	pǝdrǝ	pǝdre
ÊTRE	etjǝ	etje	sǝrǝ	sǝre
TENIR	tǝnjǝ	tǝnje	tǝdrǝ	tǝdre
SENTIR	sǝtjǝ	sǝtje	sǝtirǝ	sǝtire
VALOIR	valjǝ	valje	vodrǝ	vodre
AVOIR	avjǝ	avje	orǝ	ore

On constate que, alors que le rapport entre les formes de l'imparfait et celles du futur sont susceptibles de différer de manière largement erratique, le rapport entre les formes de première et deuxième personne à l'intérieur d'un même temps est absolument constant, pour tous les verbes du français.

Cette observation motive la postulation de deux cases distinctes dans l'espace thématique des verbes français, servant à former respectivement l'imparfait et le futur. L'idée est qu'il existe une règle par défaut qui relie ces deux cases (remplacer /j/ par /r/ pour former le FUTUR à partir de l'IMPARFAIT) mais que cette règle par défaut peut être violée par un verbe qui liste lexicalement deux thèmes distincts dans ces deux cases. Dans Bonami & Boyé (2003), nous généralisons cette idée à l'ensemble du paradigme des verbes français, et aboutissons au fait que 12 cases doivent être distinguées dans l'espace thématique d'un verbe pour rendre compte de l'ensemble des phénomènes de supplétion. L'exemple (5) précise quelles formes chaque case de l'espace thématique permet de construire.

(5)

Indice	Formes
1	IMPARFAIT, PRÉSENT 1PL et 2PL
2	PRÉSENT 3PL
3	PRÉSENT SG
4	PARTICIPE PRÉSENT
5	IMPÉRATIF 2SG
6	IMPÉRATIF 1PL et 2PL
7	SUBJONCTIF SG et 3PL
8	SUBJONCTIF 1PL et 2PL
9	INFINITIF
10	FUTUR, CONDITIONNEL
11	PASSÉ SIMPLE, SUBJ IMPARFAIT
12	PARTICIPE PASSÉ

De manière intéressante, la notion d'espace thématique peut être motivée simplement en utilisant la mesure de fiabilité utilisée par le MGL. Supposons que l'on demande au MGL de dériver la forme de l'IMPARFAIT 2PL à partir de l'IMPARFAIT 1PL. Il va aboutir à la règle $[\tilde{o} \rightarrow e / _ \#]$ qui a une fiabilité parfaite de 100%. Par contre, si l'on demande au MGL de dériver le FUTUR 1PL à partir de l'IMPARFAIT 1PL, on aboutira à un résultat beaucoup plus mitigé : la règle la plus fiable (72,8%), $[j \rightarrow \text{ər} / _ \tilde{o}\#]$, a de nombreuses exceptions (1742 sur 6499).

On aboutit donc à une caractérisation simple des espaces thématiques : les cases de l'espace thématique correspondent aux zones du paradigme qui peuvent être reliées par des règles catégoriques⁴ ; c'est en ce sens que pour savoir conjuguer un verbe il suffit de connaître son espace thématique.

⁴ Bonami & Boyé (2003) postulent l'existence de quelques *formes fléchies supplétives*, que les locuteurs sont supposées utiliser directement, sans s'appuyer sur l'espace thématique : les formes de PRÉSENT 1PL *sommes*, de PRÉSENT 2PL *faites*, *dites*, etc. On peut noter que l'existence de ces quelques exceptions à la régularité des espaces thématiques n'altère pas notablement la fiabilité des règles postulées : ainsi, malgré l'existence de *faites* et *dites*, la règle $[\tilde{o} \rightarrow e / _ \#]$ dérive avec une fiabilité de 99,7 % le PRÉSENT 2PL du PRÉSENT 1PL. Cette règle peut donc être considérée comme quasi-catégorique, et les formes fléchies supplétives ne remettent pas véritablement en cause la motivation des espaces thématiques via la fiabilité. De la même manière, l'existence d'opacités telles que celles qui sont discutées dans la section 2 diminue légèrement la fiabilité des règles reliant certaines paires de case, par exemple IMPARFAIT 1PL et IMPARFAIT 3PL : on a un comportement différent entre PARER (parj \tilde{o} , par ə) et PARIER (parj \tilde{o} , parj ə). Ces cas étant très minoritaires, la fiabilité reste cependant très haute.

L'utilisation d'un modèle comme le MGL permet donc de distinguer deux types de relations au sein des paradigmes : les relations entre membres d'une même zone correspondant à une case de l'espace thématique sont extrêmement prédictibles et quasiment constantes ; les relations entre membres de zones différentes correspondant à plusieurs cases de l'espace thématique sont moins prédictibles, et nécessitent la position de plusieurs règles concurrentes. Postuler des espaces thématiques revient simplement à faire abstraction du premier type de relation pour se concentrer sur le second. Dans la suite de cet article, nous nous concentrons sur la modélisation des relations entre cases de l'espace thématique, et laissons de côté les relations des thèmes aux formes.

2. LES BASES DE LA FLEXION RÉGULIÈRE

Il existe un présupposé qui fait l'objet d'un large consensus dans la littérature sur la flexion : un lexème régulier, c'est un lexème dont la flexion complète peut être déduite de manière déterministe à partir de la connaissance d'une forme phonologique unique, ce qu'on appellera pour les besoins de cet article la *base* du lexème. Le consensus s'arrête largement là, les approches concurrentes s'opposant tant dans l'analyse qu'elles proposent de la flexion irrégulière que dans l'implémentation précise qui est proposée pour la notion de *base*. Ainsi, dans la tradition de la phonologie générative que représente Plénat (1987), il est généralement supposé que l'existence d'une *base* unique vaut également pour de larges pans de la flexion irrégulière, ce que conteste par exemple Bonami & Boyé (2003). Par ailleurs, pour Plénat (1987) la base est une représentation sous-jacente abstraite, alors que pour Albright (2002b) c'est une forme de surface.

Dans cette section nous explorons une possibilité qui a été ignorée dans la littérature précédente : si la flexion de tout lexème régulier repose sur une *base* unique, il ne va

pas de soi que cette base corresponde toujours à la même zone du paradigme. Dans un cadre qui reconnaît la notion d'espace thématique, différentes cases de l'espace thématique peuvent servir de base pour différents lexèmes. Cette proposition nous permet d'éviter dans une large mesure la postulation de représentations phonologiques abstraites dans le traitement de la flexion régulière.

1.4. LES OPACITÉS CONVERSES DANS LA CONJUGAISON DU FRANÇAIS

Pour illustrer les enjeux, considérons la question de savoir quelle est la base d'un verbe régulier du français. Il est généralement supposé que c'est le radical de l'imparfait, le thème 1 au sens du tableau (5), qui a le statut de base. Cette hypothèse a cependant l'inconvénient majeur de ne pas permettre de différencier les verbes du premier groupe à thème 1 en /is/ (type TAPISSER : thème 1 /tapis/, thème 3 /tapis/) des verbes du deuxième groupe (type TAPIR: thème 1 /tapis/, thème 3 /tapi/). Si on veut maintenir que le thème 1 sert de base pour tous les verbes, il faut donc différencier ces deux classes de verbes par un autre biais. On pourrait envisager, par exemple, soit de considérer que les verbes du deuxième groupe ne sont pas réguliers, et n'ont donc pas de base unique (Bonami & Boyé, 2003) ; soit que la base des verbes du deuxième groupe porte un diacritique qui indique qu'elle est sujette à des règles spéciales. La nature du diacritique en question peut elle-même être débattue : il pourrait s'agir d'un trait de classe morphologique, d'un trait de règle (Plénat 1987) ou du marquage de la consonne finale comme consonne latente (Paradis & El Fenne, 1996).

Une alternative claire est cependant de simplement supposer qu'on a fait erreur en supposant que c'est le thème 1 qui sert de base. Dans le cas présent, le thème 9, ou

thème de l'infinitif, semble être un meilleur candidat⁵. A partir du thème de l'infinitif, on peut proposer les règles suivantes pour dériver les thèmes 1, 2 et 3 :

- (6) a. thème 9 \Rightarrow thème 1 : $[e \rightarrow \emptyset / _ \#] ; [\emptyset \rightarrow s / i _ \#]$
 b. thème 9 \Rightarrow thème 2 : $[e \rightarrow \emptyset / _ \#] ; [\emptyset \rightarrow s / i _ \#]$
 c. thème 9 \Rightarrow thème 3 : $[e \rightarrow \emptyset / _ \#] ; [i \rightarrow i / _ \#]$

On notera que (6) propose une multiplicité de règles pour dériver un thème d'un autre. On notera aussi que ces règles ne donnent de résultats corrects que pour les verbes réguliers : les verbes du troisième groupe, soit ne sont pas couverts par le champs des règles proposées (BATTRE, etc.) soit sont couverts mais pas correctement (MENTIR, etc.). Mais si on suppose avec Bonami & Boyé (2003) que les verbes irréguliers sont des verbes qui violent les relations par défaut entre cases de l'espace thématique, cette situation est exactement celle à laquelle on s'attend : dans les cas de BATTRE ou MENTIR, le locuteur apprend par cœur une information contradictoire avec les règles par défaut. Le véritable enjeu est donc de savoir si ces règles couvrent correctement l'ensemble des verbes des deux premiers groupes, supposés réguliers.

Dans ce contexte, on doit s'interroger sur les verbes comme APPUYER et AIGILLER. Ces verbes manifestent dans leurs thèmes 2 et 3 une opposition /i/ vs. /ij/ (/apɥi/ vs. /egɥij/) qui est neutralisée au thème 9 (/apɥije/ vs. /egɥije/) et au thème 1 (/apɥij/ vs. /egɥij/). Dans ce cas particulier, la solution semble être de considérer que c'est le thème 2 ou le thème 3 qui sert de base. Par exemple, si on suppose que c'est le thème 3, on peut proposer les règles en (7) :

- (7) a. thème 3 \Rightarrow thème 2 : $[X \rightarrow X / _ \#]$

⁵ On admet avec Boyé (2000) que le thème de l'infinitif est obtenu en retranchant le /r/ final : /tapi/ pour TAPIR,

b. thème 3 \Rightarrow thème 1 : $[C \rightarrow C / _ \#] ; [i \rightarrow j / _ \#]$ ⁶

c. thème 3 \Rightarrow thème 9 : $[\emptyset \rightarrow e / C _ \#] ; [\emptyset \rightarrow je / i _ \#] ; [\emptyset \rightarrow je / wa _ \#]$ ⁷

On a maintenant proposé deux jeux de règles concurrentes pour déterminer une partie de l'espace thématique des verbes à partir d'une base : les règles (6) partent du thème 9 de l'infinitif, alors que les règles (7) partent du thème 3 du présent singulier. Pour la grande majorité des verbes réguliers, qui ont une finale consonantique au thème 3 à laquelle s'ajoute un /e/ au thème 9, les deux systèmes de règles donnent exactement le même résultat. Mais pour chacun des deux jeux de règles, il existe une classe de verbes pour lesquelles elles ne donnent pas des résultats corrects : les règles (6) échouent pour les verbes comme APPUYER, les règles (7) échouent pour les verbes du deuxième groupe comme FINIR. On se trouve face à un système d'opacité converses: certaines distinctions transparentes au thème 9 sont opaques au thème 3, et inversement certaines distinctions opaques au thème 3 sont transparentes au thème 9.

Ce type de situation est typique de celles qui motivent classiquement le recours à des représentations phonologiques abstraites, qui rendent explicite ce qui est masqué par les opacités de surface. On peut cependant se demander à quel point l'utilisation de représentations abstraites est indispensable dans ce cas ; nous explorons ci-dessous une solution alternative directement inspirée d'idées issues du travail sur le *Minimal Generalization Learner*.

1.5. FONCTIONS THÉMATIQUES ET BASE VARIABLE

/tapise/ pour TAPISSER. La non-réalisation de /r/ pour les verbes du premier groupe est analysée comme l'effet d'une règle phonologique générale qui supprime /r/ après /e/ en fin de mot.

⁶ On suppose que les /i/ apparaissant après des amas consonnantiques suivis de /j/ sont insérés par une règle phonologique générale. La représentation sous-jacente du thème de *appuyons* est donc /apuj/

⁷ La troisième règle traite le cas des verbes à thème 3 en /-wa/, type PLOYER. Pour les verbes en /e/ comme CRÉER voir plus bas note 8.

Au début de cette section, on a supposé qu'un verbe régulier est un verbe dont toute la flexion se déduit de la connaissance d'une seule base ; en revanche rien ne nous oblige à penser que c'est pour tous les verbes réguliers la même case de l'espace thématique qui sert de base. On pourrait donc adopter le dispositif suivant pour rendre compte des données du français :

- (8) a. A chaque case de l'espace thématique est associée une série de *fonctions thématiques* indiquant comment chaque autre case lui est associée dans le cas régulier⁸ ; (6) et (7) sont des spécifications de certaines des fonctions thématiques partant du thème 9 et du thème 3 respectivement.
- b. Pour un lexème donné, chaque case de l'espace thématique peut être soit remplie lexicalement, soit déduite à l'aide des fonctions thématiques ; la spécification lexicale a priorité sur l'utilisation des fonctions thématiques.
- c. Un verbe régulier est un verbe pour lequel une seule case de l'espace thématique est remplie lexicalement. En revanche l'identité de la case concernée peut varier d'un verbe régulier à l'autre.

Les fonctions thématiques sont des fonctions partielles, parce qu'il existe crucialement des cas sur lesquels elles ne disent rien. Par exemple, la fonction spécifiée en (6a) ne dit rien des thèmes d'infinitif à finale consonantique, qui correspondent tous à des verbes du troisième groupe, pour lesquels on sait que le rapport entre thème 9 et thème 1 n'est pas prédictible. La spécification de ces fonctions prend la forme d'une série de règles phonologiques, mais il est crucial que ces règles soient exclusives, pour que la relation entre cases de l'espace thématique reste une relation fonctionnelle. Autrement dit, à ce

⁸ Nous supposons ici pour simplifier que chaque case de l'espace thématique est relié à chaque autre case par une fonction thématique, ce qui implique l'existence de $12 \times 11 = 132$ fonctions thématiques. Si on suppose, comme Bonami & Boyé (2003), que l'espace thématique est structuré comme un *arbre de dépendances thématiques*, le nombre de fonctions qu'il est nécessaire de postuler diminue drastiquement : seules les cases qui sont adjacentes dans l'arbre de dépendance doivent être reliées (ce qui aboutit à 22 fonctions thématiques avec l'arbre de Bonami & Boyé,

niveau de modélisation, il n'est pas question de mettre en concurrence des règles applicables au même input, comme dans le MGL.

Si on adopte les hypothèses en (8), on peut maintenant proposer un système général pour la flexion des verbes réguliers au présent et à l'infinitif, donné ici en (9)⁹.

(9)

(9) a. *thème 2* \Rightarrow *thème 1* :

i) $\emptyset \rightarrow \emptyset / [-\text{syllabique}] _ \#$ iii) $\emptyset \rightarrow \emptyset / e _ \#$

ii) $\emptyset \rightarrow j / wa _ \#$

iv) $\left[\begin{array}{l} -\text{consonantique} \\ +\text{vocalique} \\ +\text{haut} \end{array} \right] \rightarrow [-\text{vocalique}] / _ \#$

b. *thème 2* \Rightarrow *thème 3* :

$\emptyset \rightarrow \emptyset / [+ \text{syllabique}] _ \#$

c. *thème 2* \Rightarrow *thème 9* :

$\left[\begin{array}{l} -\text{consonantique} \\ +\text{vocalique} \\ +\text{haut} \end{array} \right] \rightarrow [-\text{vocalique}] e / _ \#$

d. *thème 3* \Rightarrow *thème 2* :

$\emptyset \rightarrow \emptyset / [-\text{syllabique}] _ \#$

e. *thème 3* \Rightarrow *thème 9* :

i) $\emptyset \rightarrow je / wa _ \#$

ii) $\emptyset \rightarrow e / [-\text{syllabique}] _ \#$

2003). Des hypothèses intermédiaires mériteraient également d'être explorées : nous laissons pour des recherches futures l'examen de l'interaction entre les fonctions thématiques et la structuration de l'espace thématique.

⁹ Le lecteur attentif notera que quelques règles ont été reformulées depuis (6-7). Ces changements permettent de s'assurer que pour *tous* les verbes des deux premiers groupes, toutes les règles s'appliquent correctement, sans contradiction. Nous remercions Aurélien Giraud pour sa contribution à la formulation finale de l'analyse explicitée en (9-10).

f. *thème 9* \Rightarrow *thème 2* :

i) $\emptyset \rightarrow s / i \text{ ___ } \#$

iii) $e \rightarrow \emptyset / [+consonnantique] \text{ ___ } \#$

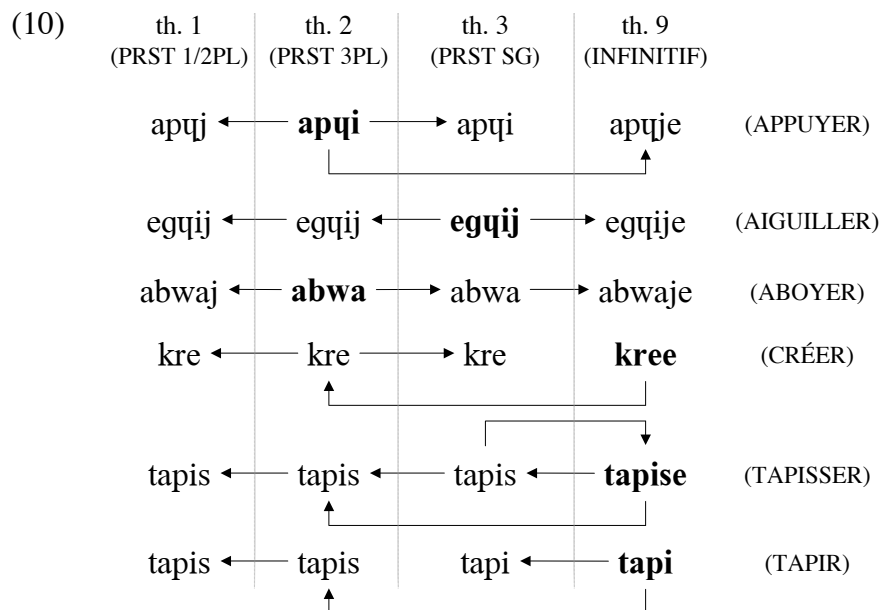
ii) $e \rightarrow \emptyset / \left[\begin{array}{l} -\text{consonnantique} \\ +\text{vocalique} \end{array} \right] \text{ ___ } \#$

g. *thème 9* \Rightarrow *thème 3* :

i) $e \rightarrow \emptyset / [+consonnantique] \text{ ___ } \#$

ii) $\emptyset \rightarrow \emptyset / i \text{ ___ } \#$

L'exemple (10) indique quel thème sert de base pour chacun des verbes que nous avons discuté (en gras). Chaque flèche indique que la fonction thématique correspondante est définie pour la forme de départ en question et pointe vers le résultat de son application.



L'utilisation de relations mutuelles entre cases de l'espace thématique permet donc de sauver l'hypothèse de l'unicité de la base pour les verbes réguliers, tout en évitant le recours à des représentations phonologiques abstraites ou à des diacritiques arbitraires.

Les quelques exemples présentés ici ne couvrent bien sûr pas la totalité de l'espace thématique : une spécification complète de la flexion du français doit préciser pour chaque case comment elle est reliée aux onze autres. De plus nous n'avons pas tenu compte de la totalité des phénomènes de neutralisation qui demandent à être modélisés¹⁰.

Une caractéristique du système de fonctions thématiques proposé ici est qu'il est partiellement redondant : on suppose que les locuteurs possèdent dans beaucoup de cas des règles permettant de déduire aussi bien le thème X du thème Y que le thème Y du thème X. Si cette redondance partielle va à l'encontre des habitudes dans les modèles de compétence, elle est à la fois motivée par la modélisation des opacités converses et par les données de performance de base : en dehors de ces cas d'opacité, les locuteurs sont capables de conjuguer un nouveau verbe régulier sur la base d'une quelconque de ses formes ; par exemple face à une nouvelle forme de présent *il clarve* le locuteur sera capable de déduire un infinitif *clarver*, aussi facilement qu'à l'inverse il pourra déduire le présent de l'infinitif. Il n'y a que dans les cas d'opacité manifeste (par exemple face à un infinitif comme *radriller*) que le locuteur n'est pas à même de faire un choix catégorique. Notre hypothèse consiste donc simplement à inclure dans le modèle de compétence des relations mutuelles dont l'existence ne fait aucun doute au niveau de la performance.

Pour conclure, il est utile de comparer le modèle esquissé ici avec le MGL de Albright & Hayes. L'idée de prendre en compte dans la modélisation de la flexion des

¹⁰ Un autre cas à prendre en compte est celui de l'opposition entre RAMENER, PEINER et DÉJEUNER : RAMENER et PEINER ont des thèmes 3 indistinguables (/ramɛn/ vs. /pɛn/) mais se distinguent par leur thème 9 (/raməne/ vs. /pene/), alors qu'à l'inverse RAMENER et DÉJEUNER sont indistinguables au thème 9 — du moins pour les locuteurs qui traitent le /æ/ de DÉJEUNER comme un schwa instable — (/raməne/ vs. /deʒəne/) mais pas au thème 3 (/ramɛn/ vs. /deʒɛn/). On peut rendre compte de cette situation en ajoutant à (6c) la règle [əCe → ɛC / __ #], et en posant les spécifications lexicales en (i) :

(i) a. RAMENER : thème 9, /raməne/
 b. PEINER : thème 9, /pene/

collections de règles phonologiques reliant deux formes est directement inspirée du MGL ; de plus l'utilisation de relations multilatérales entre cases de l'espace thématique est inspirée des réflexions d'(Albright, 2002b) sur l'identification des bases dans les paradigmes. Le modèle résultant a cependant des propriétés très différentes. D'une part, le modèle esquissé ici est nettement un modèle de compétence, dans lequel nous ne cherchons pas à modéliser la flexion irrégulière, qui est supposée reposer sur le listage lexical de formes idiosyncrasiques. A l'inverse le MGL vise à modéliser le comportement des locuteurs précisément dans les situations où ils tentent de prédire une forme pour un lexème irrégulier. En conséquence, les collections de règles utilisées ici sont très différentes de celles du MGL : les fonctions thématiques sont définies sur la base de collections de règles phonologiques exclusives, qui partitionnent l'ensemble des thèmes donnés en entrée; ces règles ont toutes une fiabilité haute, puisqu'elles n'ont pas d'exception parmi les lexèmes réguliers. Dans le MGL, il est crucial au contraire que plusieurs règles non-exclusives, et qui contrastent en termes de fiabilité, soient considérées.

D'autre part, nous avons défendu ici la thèse selon laquelle ce n'est pas toujours la même case de l'espace thématique qui sert de base à la flexion régulière. Cette thèse va directement à l'encontre de la prémisse d'(Albright, 2002b), selon laquelle il existe une case unique du paradigme qui sert de base, dont l'identification est une des tâches essentielles de l'apprenant.

3. RÉFÉRENCES

- Albright, Adam., 2002a., “Islands of Reliability for Regular Morphology: Evidence from Italian”, *Language* n°78, p. 684-709.
- Albright, Adam, 2002b, *The identification of bases in morphological paradigms.*, thèse de doctorat, UCLA.
- Albright, Adam et Bruce Hayes. 2002. Modeling English past tense intuitions with Minimal Generalization, dans M. Maxwell, *Proceedings of the 6th meeting of the ACL Special Interest Group in Computational Phonology*.
- Albright, Adam et Bruce Hayes, 2003, “Rules vs. analogy in English past tenses: a computational/experimental study”, *Cognition* n°90, p. 119-161.
- Aronoff, Mark, 1994, *Morphology by itself*, MIT Press, Cambridge.
- Bonami, Olivier et Gilles Boyé, 2003, “Supplétion et classes flexionnelles dans la conjugaison du français”, *Langages* n°152, p. 102-126.
- Bonami, Olivier et Gilles Boyé, sous presse, “Construire le paradigme d’un adjectif”. *Revue de Linguistique de Vincennes*.
- Bonami, Olivier, Gilles Boyé et Françoise Kerleroux, à paraître, “L’allomorphie radicale et la relation flexion-construction”, dans B. Fradin, F. Kerleroux et M. Plénat, *Aperçus de morphologie*.
- Boyé, Gilles, 2000, *Problèmes de morpho-phonologie verbale en français, espagnol et italien*, thèse de doctorat, Université Paris 7.
- Chomsky, Noam et Morris Halle, 1968, *The Sound Pattern of English*, Harper & Row, New York.
- de Calmès, M. and G. Pérennou. 1998. “Bdlex : a lexicon for spoken and written french.”, dans *1st International Conference on Language Resources and Evaluation*, pages 1129–1136. Grenade: ELRA.

Paradis, Carole et Fatimazohra El Fenne, 1995, "French verbal inflection revisited : Constraints, repairs and floating consonants", *Lingua* n°95, p. 169-204.

Plénat, Marc, 1987, "Morphologie du passé simple et du passé composé des verbes de l' "autre" conjugaison", *ITL Review of Applied Linguistics* n°77, p. 93-150.

Schane, Sanford, 1968, *French Phonology and Morphology*, MIT Press, Cambridge.