

Lexicographie, terminologie, dictionnairique

21 septembre 2021

—Kata Gábor

kata.gabor@inalco.fr—

Lecture pour la semaine prochaine :

A. Kilgarriff : *Word Senses*. In *Word Sense Disambiguation : Algorithms and Application*, ed. Agirre and Edmonds. Chapter 2, pp 29-46. (disponible sur <https://www.kilgarriff.co.uk/publications.htm>)

Lexicographie et information lexicale

- *arbitraire du signe* linguistique : aucun rapport entre la forme du signe (le mot) et le signifié (le concept qu'il désigne)
- grammaire vs lexicque : le vocabulaire est arbitraire, c'est la grammaire qui contient les régularités (la façon dont les mots sont organisés en phrases)
- au sein de la linguistique, la lexicologie est l'étude des mots et leurs propriétés, les informations *lexicales*

Le sujet de la lexicologie

- mot : unité forme + sens
- lexicologie : étude des mots, de leur sens, et de leurs propriétés lexicales
- lexicographie : pratique de rédiger, construire des dictionnaires (choix des vedettes, descriptions de sens, définitions etc)
- lexicologie computationnelle :
 1. études lexicologiques assistées par ordinateur
 2. création de ressources lexicologiques (dictionnaires) pour le TAL

Produits lexicographiques

1. produits lexicographiques pour usage humain

- monolingues
 - usage normatif vs descriptif
 - explanatoires, encyclopédique
 - spécialisés : domaines spéciaux (vocabulaire spécifique, terminologie), registre spécial (argot etc)
 - spécialisés : informations spécifiques (étymologique, dialectal)
- bilingues
 - langue source (langue des mots vedettes) et langue cible (les traductions)
 - encodage (langue source : langue maternelle)
 - décodage (langue source : langue étrangère)

2. produits lexicographiques computationnels

- lexique de base (liste des mots ; liste des formes)
- lexique morphosyntaxique (mots+categories)
- lexique sémantique
- ressources structurées, ontologies (contiennent des relations)
- ressources spécialisées (terminologie, mots clés, topic models pour l'indexation de documents)
- lexiques bilingues pour la traduction automatique
- représentations vectorielles et *word embeddings* (décomposition du sens)

Les théories du sens et les approches lexicologiques/lexicographiques

objectifs principaux de la lexicographie

1. distinguer et énumérer le sens des mots
2. décrire le sens

Un bon inventaire de sens doit permettre à l'utilisateur de sélectionner le sens qui correspond à une occurrence du mot en contexte (Word Sense Disambiguation, WSD)

Les approches lexicographiques et leurs théories

1. Lexicologie traditionnelle : approche rationaliste

- bases 'platoniques' (les idées sont des entités objectives existantes)
- signifié - signifiant (Saussure); sens et référence (Frege)
- le sens d'un mot n'est pas égal à sa référence, mais existe comme réalité objective et permet de faire référence à un concept
- approche introspective
 - il existe un inventaire des sens des mots qui est objectif (inter-subjectif), chaque mot w a un nombre n de sens bien distincts,
 - les sens sont mutuellement exclusifs et bien délimités.
- dans la pratique lexicologique, cela soulève des problèmes :
 - les lexicographes ne sont pas toujours d'accord sur les frontières et les distinctions fines (*lumping or splitting*),
 - les locuteurs ne sont pas toujours d'accord sur le sens d'un mot dans un contexte donné,
 - il existe des phénomènes d'extension de sens (usages plus ou moins figurés, métaphores)

2. Lexicologie moderne : approche empirique

- on parle d'usages d'un mot plutôt que de sens
- on ne présuppose pas l'existence d'un inventaire de sens objectif
- pratique : observer les différentes utilisations du mot en contexte (phrases dans un corpus)
- hypothèse distributionnelle : les mots ayant un sens similaire apparaissent dans des contextes similaires
 - (a) lexicologie assistée par le corpus, **corpus-based lexicology**
 - les occurrences dans le corpus aident à délimiter les sens
 - parallèlement à l'identification des sens, il faut observer et distinguer les 'patterns' (motifs) d'utilisation
 - (b) lexicologie guidée par le corpus, **corpus-driven lexicography**
 - les sens émergent du corpus en regroupant les contextes
 - l'hypothèse distributionnelle est exploitée en tant que "définition" : les contextes d'utilisation = le sens
 - la différence est plutôt méthodologique : aucun a priori humain dans les approches guidées par corpus

Polysémie, homonymie et sens vagues

1. homonymie : plusieurs sens qui ne sont pas liés (avocat)
 - catégorie différente (est)
 - catégorie identique → un mot ou plusieurs ?
2. polysémie : plusieurs sens liés (tête)
3. sens vague/sous-spécifié (oncle : frère de la mère ou frère du père; noms de type : fleur, fruit,...; verbes : partir, ...)

comment les distinguer ?

- critère logique : X et non X
 - polysémie : C'est un étrier (os), et pas un étrier (de cheval).
 - homonymie : C'est un avocat (fruit), pas un avocat (personne).
 - sens vague : ?? C'est mon oncle et pas mon oncle.
- coordination et ellipse
 - polysémie : * Pierre a vu l'étrier (os), et Jean aussi (a vu l'étrier (de cheval)).
 - homonymie : *Pierre a vu l'avocat (fruit), et Jean aussi (a vu l'avocat (personne)).
 - sens vague : Pierre a vu son oncle, et Jean aussi (a vu son oncle (à lui)).
- critère de définition

peindre,n

- Recouvrir une surface, un objet d'une couche de peinture : Peindre les murs en blanc.
- Représenter quelque chose à l'aide de peinture sur un support : Peindre un numéro sur un écriteau.
- Orner une surface de couleurs, de motifs avec une peinture ou une autre matière : Peindre de la porcelaine.
- Réaliser en peinture un tableau, y représenter quelque chose à la peinture : Il ne peint que des portraits.
- Décrire, représenter quelque chose ou quelqu'un, par des mots, dans un texte, un discours : Romancier qui a peint la vie à la campagne.