

MORFETIK, UN DICTIONNAIRE MORPHOLOGIQUE : ILLUSTRATION AVEC LE LEXIQUE DE LA PERCEPTION

Aude GREZKA

Université Paris 13, Sorbonne Paris Cité & CNRS (France)

UMR 7187 Laboratoire Lexiques Dictionnaires Informatique

Abstract (En): Presently, numerous works concern the reflection and the conception of appropriate methodologies for automatic language processing. Their objectives are to create a linguistic automatable resource which will be used on IT applications (assistant to writing, to translation, etc.). In this article, we present the morphological resource Morfetik, a morphological dictionary of the simple words of French. At first, we return to the difference between a computer dictionary and computerized dictionary. Secondly, we approach the structuring of Morfetik by illustrating it with the lexicon of the perception.

Keywords (En): morphological dictionary ; flexion ; lemma ; lexicon ; perception

Mots-clés (Fr) : dictionnaire morphologique ; flexion ; lemme ; lexique ; perception

Introduction

De nombreux travaux portent actuellement sur la réflexion et la conception de méthodologies appropriées au traitement automatique des langues. Il s'agit de modéliser une certaine partie de la connaissance linguistique afin de la rendre exploitable par la machine, de créer de la ressource linguistique qui sera employée dans des applications informatiques (aide à la rédaction, à la traduction, etc.). Dans ce travail, nous présentons la ressource morphologique Morfetik, un dictionnaire morphologique des mots simples du français. Dans un premier temps, nous revenons sur la différence entre dictionnaire informatique et dictionnaire informatisé. Dans un deuxième temps, nous abordons la structuration de Morfetik en l'illustrant avec le lexique de la perception.

1. Problématique et axes de recherche

Le traitement automatique des langues nécessite une identification précise des formes, une conceptualisation des phénomènes linguistiques. Cela suppose un recensement lexical complet et un degré suffisant de couverture, que peut apporter un « dictionnaire électronique » construit à cet effet.

Il est important de revenir sur la différence entre deux types de dictionnaires qui utilisent l'informatique : dictionnaire *informatisé* (ou *numérique*) et dictionnaire *électronique* (COURTOIS et SILBERZTEIN, 1990 ; GROSS, 2005).

Le dictionnaire *informatisé* représente la version électronique du dictionnaire-papier et a l'avantage que la recherche de l'information y est infiniment plus rapide puisqu'il y a une véritable navigation dans le texte. Mais les informations que l'on est en mesure de mettre en évidence ne sont pas de nature différente de celles que l'on trouve dans un dictionnaire-papier. Ce type de dictionnaire a une norme de codage inhérente au phénomène de rétroconversion.

À ce dictionnaire informatisé, on opposera le dictionnaire *électronique*, qui est conçu dans un but de traitement automatique (reconnaissance ou génération). Toutes

les informations y sont explicites et reproductibles. Puisqu'il a pour objectif le traitement automatique, il ne s'appuie pas sur des rectifications analogiques qu'un lecteur humain est en mesure de faire quand une information est incomplète. Ce type de dictionnaire ne présuppose pas une connaissance préalable de la langue et du monde. Seules les informations qui lui sont fournies et sa structuration permettent de reconnaître ou de construire les phrases d'une langue naturelle. Un dictionnaire électronique est une base de données cohérente, qui n'existe que sur support électronique et qui n'a donc pas subi d'informatisation. Il sera directement utilisable par l'ordinateur.

Observons comment est traité linguistiquement un verbe de perception comme *voir* dans un dictionnaire informatisé monolingue du type Le Robert :

voir [vwɔʁ] v. <conjug. : 30>

• XII^e *veoir*; *vedeir* 980; *lat.* *videre*

I ♦ V. intr. (1080 *vedeir*) Percevoir les images des objets par le sens de la vue. C'est « un postulat bien ancré, qu'un nouveau-né [...] "ça" ne voit pas » (F. Leboyer). Ne plus voir : perdre la vue (⇒ *aveugle*). Ne voir que d'un œil (⇒ *borgne*). Regarder sans voir. Voir trouble, confusément, mal, à peine (cf. fam. Ne pas avoir les yeux en face des trous). On ne voit pas à dix pas. « Tu as bu, Grémio, tu vois double » (Musset). Voir clairement, distinctement. On commence à y voir clair. Il n'y voit pas très bien : il souffre d'un trouble, d'une affection de la vue (⇒ fam. *bigleux*, *mîro*). N'y voir goutte*. « Tandis qu'à peine à tes pieds tu peux voir » (La Fontaine). Mettez vos lunettes pour mieux voir. Voir au loin, très loin (cf. Avoir des yeux de lynx*, d'aigle*).

◊ Loc. Voir loin : prévoir. Ne pas voir plus loin que le bout de son nez. — Ne voir que par les yeux de qqn, se fier entièrement à son jugement, suivre son opinion en tout. — Voir avec les yeux de la foi : considérer les choses à la lumière de la foi.

II ♦ V. tr. dir. (980)

1 ♦ Percevoir (qqch.) par les yeux. Voir qqch. de ses yeux, de ses propres yeux. « Je le vis, je rougis, je pâlis à sa vue » (Racine). Allus. littér. J'ai vu, de mes yeux vu. Les aventuriers « avaient vu, de leurs yeux vu la mer Pacifique » (Michelet). Je le vois très bien. ⇒ *distinguer*. Il a tout vu, tout observé sans être vu. Je l'ai à peine vu. ⇒ *apercevoir*, *entrevoir*. « Voir tout cela d'un clin d'œil » (Balzac). ⇒ *embrasser*, *saisir*. « Je faisais semblant de ne pas les voir » (France). « Dans la haute montagne, il avait vu de près des glaciers » (Aragon). Que vois-je ? (exprime l'étonnement). Une femme agréable à voir, plaisante, jolie. Cela fait plaisir à voir. Demander à voir les cartes d'un joueur. Payer pour voir (au poker). Fig. Payer par curiosité. — J'ai vu cela dans le journal. ⇒ *1. lire*.

◊ Loc. Voir le jour : naître; paraître (choses). N'y voir que du feu. En voir trente-six chandelles. Loc. Voir la paille dans l'œil du voisin et ne pas voir la poutre dans le sien. Je l'ai vu comme je vous vois, aussi réellement. — FAIRE VOIR. ⇒ *montrer*. Faites-moi voir les lieux, le fonctionnement de l'appareil. Il nous a fait voir ce qu'il fallait faire, comment il fallait procéder. — Se faire voir : se montrer (personnes). Il essaie de ne pas se faire voir : il se cache. « Il y a des endroits où il faut se faire voir » (La Bruyère). Se faire voir avec qqn. Se faire voir sous son bon jour. Loc. fam. Va te faire voir (chez les Grecs) ! (cf. Aller au diable).

Figure 1 : Extrait de l'entrée *voir* (CD Rom Le Petit Robert)

Les données telles qu'elles se trouvent dans la définition présentent de nombreux inconvénients et faiblesses (CATACH, 2000). Nous avons un enchevêtrement d'informations de différentes natures (définitions, exemples, expressions, etc.), des illogismes de classification de sens et de synonymie, des renvois hypertextes non ciblés, etc. Si l'on prend simplement l'exemple des locutions au sein de la définition (*voir loin*, *voir le jour*, *se faire voir avec quelqu'un*, etc.), on constate une description lexicographique assez pauvre. Les informations d'emplois de ces locutions sont pratiquement inexistantes. De façon assez surprenante également, on trouve dans le sens 1 de l'emploi transitif direct (« percevoir quelque chose par les yeux ») des locutions comme *voir le jour* ou *se faire voir* qui sont sémantiquement très éloignées de ce sens.

Nous voyons que les technologies informatiques, dans le cadre des dictionnaires informatisés, si elles ont transformé le mode d'accès aux données contenues dans

ces documents¹, ne semblent cependant pas avoir été exploitées au maximum par les lexicographes et les éditeurs. Ces améliorations n'ont qu'un impact limité sur l'exploitation des richesses du dictionnaire dans le cadre de l'apprentissage d'une langue. Ainsi, du point de vue de la description des propriétés linguistiques des unités lexicales, un important travail reste à faire. L'informatisation des dictionnaires papier, bien que constituant une première approche dans le traitement du langage par ordinateur (puisqu'elle génère des outils uniquement consultables sur ordinateur), ne répond pas à une véritable description formelle du langage².

Aux dictionnaires informatisés, mis à la disposition de l'utilisateur humain doté d'une connaissance linguistique, s'opposent les dictionnaires électroniques, les bases de données, uniquement élaborés pour être utilisés par des programmes informatiques... et dépourvus, normalement, de connaissances linguistiques. Face au caractère inachevé des dictionnaires traditionnels et à leur grande part d'implicite apparaît la nécessaire complétude des dictionnaires électroniques susceptibles d'analyser les mots d'un texte et le caractère obligatoirement explicite de leurs informations. Enfin, l'association étroite des dictionnaires électroniques à des programmes de traitement automatique impose une cohérence structurale des entrées. L'émergence de ces dictionnaires s'inscrit dans la mouvance du traitement automatique du langage (TAL), discipline visant à traiter de façon automatique, c'est-à-dire par le biais de l'outil informatique, les données linguistiques de tout genre (COUGNON et FAIRON, 2009).

2. Morfetik

Les dictionnaires morphologiques, entre autres, sont utiles pour l'analyse automatique des textes. Dans cette perspective, nous avons élaboré la ressource linguistique Morfetik. Cette ressource est le résultat du travail d'une vingtaine d'années de collecte et de description (MATHIEU-COLAS, 2009 ; GREZKA, CARTIER et MATHIEU-COLAS, 2015)³. Morfetik est un dictionnaire morphologique des mots simples du français. Le système ainsi conçu permet de générer automatiquement l'ensemble des formes simples du français – 758 035 formes fléchies pour 103 312 lemmes (compte tenu des homographies) – en l'état actuel de la description : 66 393 noms, 10 223 verbes et 24 391 adjectifs (2 305 autres). L'inventaire n'est pas clos. Actuellement, nous recensons l'ensemble des rectifications orthographiques du français de 1990 (trait d'union remplacé par la soudure, accent circonflexe, etc.).

Cette ressource permet d'obtenir, pour n'importe quel mot français, l'ensemble de ses formes (pluriel des noms, féminin et pluriel des adjectifs, formes conjuguées

¹ Les versions électroniques des dictionnaires ont simplifié l'accès aux informations contenues dans le dictionnaire par la mise en place de fonctionnalités de recherche plein texte et paramétrables et par les améliorations apportées à la lisibilité de l'article (mise en page, mise en évidence de la structure de l'article, etc.).

² Dans la pratique didactique courante, le dictionnaire informatisé s'utilise essentiellement comme ouvrage d'appoint dans des situations de communication quotidiennes d'aide à la compréhension ou d'aide à la rédaction.

³ Travail sous la direction à l'époque de M. Mathieu-Colas. Depuis 2014, nouvelle équipe (CNRS, universitaires, doctorant) dirigée par A. Grezka : J.F. Sablayrolles ; E. Cartier ; C. Jacquet-Pfau ; F. Martin-Berthet ; M. Mathieu-Colas ; W. Dekdouk ; L. Catach.

des verbes), ou bien, réciproquement, d'identifier le mot (la forme de base, le "lemme") correspondant à n'importe quelle forme fléchie. Un ensemble d'outils de navigation et de consultation de la ressource a ainsi été développé :

- outil de lemmatisation de formes : à partir d'une forme quelconque, l'outil renvoie à l'ensemble des lemmes pouvant y correspondre ;
- outil de génération des flexions : à partir d'une zone permettant de saisir une forme, l'outil fournit les différents lemmes possibles, puis les différentes formes liées à chaque lemme ;
- moteur de recherche sur les formes et/ou les flexions : un moteur de recherche permet de lancer des recherches classiques sur l'ensemble des champs disponibles dans la base. L'idée est de proposer aux chercheurs un moteur de recherche permettant d'effectuer des recherches complexes sur la base lexicographique.

La ressource peut soit être consultée via un moteur de recherche, soit téléchargée dans un format XML⁴.

3. Illustration de Morfetik avec les verbes de perception : structuration interne

Pour notre test et notre démonstration, nous avons utilisé le matériel lexical sur la perception de GREZKA A. (2006), soit :

MATERIEL LEXICAL	NOMBRE D'ENTREES
Verbes de perception	129
Verbes d'émission de phénomènes sensibles (visuel, sonore, olfactif)	424
Adjectifs de sensation	1142
Noms relatifs aux phénomènes auditifs, olfactifs et gustatifs (bruit, odeur, goût)	2903
Noms relatifs aux instruments de perception	123

En interne (c'est-à-dire la structuration non visible par l'utilisateur), les données de Morfetik⁵ sont structurées sous forme de tables et servent de point de départ à un système de traitement qui combine un moteur de flexion, un dictionnaire des formes fléchies, des interfaces de consultation et d'interrogation, ainsi qu'un ensemble d'outils permettant la maintenance et l'exploitation des ressources. La structure des tables diffère selon les catégories morphosyntaxiques : 5 groupes distincts ont été constitués (adverbes, déterminants/pronoms, noms, adjectifs et verbes).

⁴ A partir du site du laboratoire LDI : <https://outilsldi.univ-paris13.fr/morfetik/>

⁵ Les données ont été établies à partir de nombreuses sources lexicographiques, notamment : le *DELAS* (Dictionnaire électronique du LADL, cf. B. COURTOIS 1990) ; le *Petit et le Grand Robert* ; le *Petit Larousse illustré*, le *Lexis*, le *Grand Larousse encyclopédique* et le *Grand Dictionnaire encyclopédique Larousse* (GDEL) ; le *Trésor de la langue française* ; le *Harrap's* et le *Robert & Collins* ; des dictionnaires d'argot ; des tables de conjugaison (dont le *Bescherelle* et les *Verbes logiques* de A. DUGAS) ; *Le Bon Usage* de GREVISSE et des dictionnaires de « difficultés » pour le traitement des cas problématiques.

Ainsi, pour les mots invariables, comme les adverbes, un listage est suffisant. En revanche, pour les déterminants et les pronoms, les lemmes sont accompagnés de l'énumération des formes ; par exemple, pour *un* :

Lemme	CatGram	Forme	Genre	Nombre
un	D:Ind	un	M	S
un	D:Ind	une	F	S
un	D:Ind	des	M	P
un	D:Ind	des	F	P
un	D:Num	un		
un	P:Ind	un	M	S
un	P:Ind	une	F	S
un	P:Ind	uns	M	P
un	P:Ind	unes	F	P

Figure 2 : Extrait de la table des déterminants

Pour les mots les plus complexes (noms, adjectifs et verbes), deux tables complémentaires ont été réalisées :

- (i) des tables de flexion permettant d'identifier et de coder tous les types de formation (pluriels, conjugaisons, etc.)
- (ii) des tables attribuant à chaque lemme le code flexionnel correspondant (utilisées ensuite par le moteur de flexion pour produire l'ensemble de toutes les formes fléchies).

Ce sont ces tables qui seront ensuite utilisées par le moteur de flexion pour produire l'ensemble de toutes les formes fléchies.

Pour les noms, la table des codes comporte 63 codes. Outre les pluriels courants, toutes les formes particulières, y compris les nombreux types d'emprunts, reçoivent un code spécifique. Pour des raisons à la fois théoriques et pratiques, les noms n'ont pas été fléchis en genre (il n'y a donc dans la nomenclature que des noms masculins et des noms féminins) :

Code	Radical	S	P
00	0		
01	0		s
02	0		x
03	1	l	ux
06	4	œil	yeux
1D	2	us	i

exemple (nm)	exemple (nf)
➤ sens ; albinos	albinos
➤ toucher	vue
➤ cheveu	peau
➤ bilatéral	
➤ œil	
➤ oculus	

Figure 3 : Extrait des codes flexionnels des noms : les noms de perception

Dans l'extrait, le radical correspond au nombre de caractères à enlever pour construire un radical artificiel utilisé par le fléchisseur pour générer les formes fléchies. Par exemple, pour le mot *œil* qui a un pluriel multiple, nous aurons 2 codes :

- le code 01 (*œil/œils*), pour certains emplois pluriels du mot *œil*, notamment dans les noms composés (*des œils-de-bœuf*, *œils-de-chat*, *œils-de-perdrix*, etc.). Ce sont des emplois plus rares de *œils*, que l'on va surtout rencontrer dans des domaines spécialisés comme l'armement⁶, l'imprimerie⁷, la marine⁸, l'outillage⁹, la technique¹⁰ ;
- le code 06 (*œil*, *yeux*) : *des yeux bleus*, *les yeux d'une pomme de terre*, *d'une plante*, etc.

Pour les adjectifs, 59 codes de flexion ont été nécessaires. Contrairement aux noms, les adjectifs sont plus complexes puisqu'ils sont fléchis en genre (en général, une possibilité de quatre formes différentes) :

code	rad	MS	MP	FS	FP
34	1	x	x	se	ses
40	0		s		s
42	0		s	e	es
82	1	l	ux	le	les

⁶ Œils : ouverture ménagée dans un obus pour y introduire la charge d'éclatement et visser la fusée.

⁷ Œils : partie du caractère représentant le dessin de la lettre reproduit à l'impression sur le papier pendant le tirage.

⁸ Œils : boucle formée à l'extrémité d'un filin.

⁹ Œils : trou pratiqué dans la tête d'un marteau pour y fixer le manche.

¹⁰ Œils : judas optique.

exemples
➤ harmonieux ; mélodieux ; silencieux
➤ cacophonique ; rude ; insonore
➤ criard ; déplaisant
➤ musical

Figure 4 : Extrait des codes flexionnels des adjectifs : adjectifs de sensation

Enfin, pour les verbes, la table comprend 226 codes de flexion :

CHAMPS	EXEMPLE
Numéro de code	063
Radical	-3
Radical-modèle	V-
Désinence	-OIR
Désinences des formes conjuguées (45 champs)	-ois, -ois, -oit, -oyons, oyez, -oient...
Désinences des participes (5 champs)	-oyant, -u, -ue, -us, -ues

Figure 5 : Extrait avec le verbe *voir*

Chacun des codes associés aux lemmes correspond à un type de conjugaison spécifique qui doit être défini avec précision. Les « radicaux » et « désinences » ainsi décrits ne correspondent pas nécessairement au découpage morphologique traditionnel¹¹. Afin de faciliter le traitement automatique, le radical est défini comme le plus petit dénominateur commun : par exemple, ici, V- pour VOIR. A la limite, le radical peut être une forme vide (*être*, *avoir*, *aller*). De manière complémentaire, les désinences (-oir, -ois, -oit, -oyons, etc.) s'ajoutent au radical pour construire les formes fléchies.

Ainsi, à partir de cette double représentation, le moteur de flexion peut produire l'ensemble des formes fléchies. Il peut également être exploité pour reconnaître des formes inconnues. La ressource permet ainsi d'avoir l'ensemble des formes (pluriel des noms, féminin et pluriel des adjectifs, formes conjuguées des verbes), ou bien, réciproquement, d'identifier le mot (la forme de base, le "lemme") correspondant à n'importe quelle forme fléchie.

Par exemple, si nous tapons « vu », l'interface de la ressource nous proposera le tableau suivant, qui montre les différentes analyses possibles de la forme :

¹¹ Ce découpage morphologique est un découpage interne, donc non visible par l'utilisateur de la ressource.

Entrer une forme simple : ☐ Sensible aux accents

La réponse du système pour la forme *vu*

Lemme	Categorie	Temps	Nombre	Genre	Personne
voir	Verbe	Pp	S	M	
vu	adj		S	M	
vu	nms		S	M	
vu	Prép				

Figure 6 : Résultat d'analyse pour la forme *vu*

L'outil nous renvoie à la fois au lemme, ici à la forme nominale (*vu*), adjectivale (*vu*) et verbale (*voir*), mais également à la préposition (*vu*). Ensuite, l'utilisateur peut cliquer sur l'un des lemmes pour obtenir les formes correspondantes, et voir ainsi les différentes formes liées à chaque lemme :

vu/adj	
vu/adj	
	vu
Masculin Singulier:	<i>vu</i>
Masculin Pluriel:	<i>vus</i>
Féminin Singulier:	<i>vue</i>
Féminin Pluriel:	<i>vues</i>

vu/nms	
vu/nms	
	vu
Masculin Singulier:	<i>vu</i>

Au vu et au su de tout le monde : au grand jour
C'est du déjà vu : ce n'est pas une nouveauté.
Au vu, sur le vu de : en voyant.

vu/Verbe					
Lemme					
---	voir				
Infinitif					
---	voir				
Indicatif Présent		Indicatif Imparfait		Passé Simple	
Je	vois	Je	voyais	Je	vis
Tu	vois	Tu	voyais	Tu	vis [*]
Il	voit	Il	voyait	Il	vit
Nous	voyons	Nous	voyions	Nous	vîmes
Vous	voyez	Vous	voyiez	Vous	vîtes [*]
Ils	voient	Ils	voyaient	Ils	virent
Indicatif Futur		Conditionnel Présent		Subjonctif Présent	
Je	verrai	Je	verrais	Je	voie
Tu	verras	Tu	verrais	Tu	voies
Il	verra	Il	verrait	Il	voie
Nous	verrons	Nous	verrions	Nous	voyions
Vous	verrez	Vous	verriez	Vous	voyiez
Ils	verront	Ils	verraient	Ils	voient
Subjonctif Imparfait		Impératif Présent		Participe Présent	
Je	visse [*]			---	voyant
Tu	viesses [*]		vois	Participe Passé	
Il	vît			Masculin	vu
Nous	viussions [*]		voyons	Féminin	vue
Vous	vissiez [*]		voyez	Masculins	vus
Ils	vissent [*]			Féminins	vues

Figure 7 : Résultats de la recherche pour *vu*

Le signe ° (figure 7) indique les formes rares et/ou archaïques, notamment comme dans cet exemple avec le subjonctif imparfait (*visse*, *viesses*, *viussions*, *vissiez*, *vissent*). Il permet de formaliser les indications disséminées dans les dictionnaires (*le plus souvent*, *en général*, *surtout*, *rarement*, etc.). Linguistique, la description permet d'être plus précise et informatiquement, un tel système permet une double utilisation du lexique (standard/exhaustive), selon que les flexions marquées sont ou non prises en compte. Le choix n'est pas négligeable, compte tenu du nombre de formes en jeu. Le marquage peut intervenir à différents niveaux¹².

À ces données de base viennent s'ajouter des informations complémentaires, permettant la bonne interprétation des données. Par exemple, comme nous pouvons l'observer dans la Figure 7, avec le nom masculin singulier *vu* qu'on va rencontrer dans des expressions comme *au vu et au su de tout le monde* ; *au vu*, *sur le vu de*, etc.

¹² Sur un plan général, nous marquons :

- tous les passés simples à la 2e personne : *vis*[°], *vites*[°], *regardas*[°], *regardâtes*, ... (le marquage de la 1re pers. se justifierait moins) ;
- tous les imparfaits du subjonctif, sauf la 3e personne du singulier : *visse*[°], *viesses*[°], *viussions*[°], *vissiez*[°], *vissent*, ...

Les termes spécialisés gagnent également à être marqués comme tels, à l'exemple des noms que nous avons recueillis dans le domaine de l'ophtalmologie :

ID	Lemme	CatGramm	Flex	Dom
N61	amblyopie	nf	01	med:ophtalm.
N80	astigmatisme	nm	01	med:ophtalm.
N233	blépharite	nf	01	med:ophtalm.
N523	cataracte	nf	01	med:ophtalm.
N575	cécité	nf	01	med:ophtalm.
N1018	ectropion	nm	01	med:ophtalm.
N1019	entropion	nm	01	med:ophtalm.
N1020	glaucome	nm	01	med:ophtalm.
N1021	hypermétropie	nf	01	med:ophtalm.
N1023	kératite	nf	01	med:ophtalm.
N1024	kératocône	nm	01	med:ophtalm.
N1125	névrite optique	nf	01	med:ophtalm.
N1227	uvéite postérieure	nf	01	med:ophtalm.

Figure 8 : Résultats de la recherche par domaine, ophtalmologie

4. Perspectives

Les systèmes de traitement automatique des langues (TAL) intègrent de plus en plus des connaissances de nature linguistique. Parmi ces connaissances, la morphologie est très souvent utilisée, en particulier dans le traitement des langues comme le français, à morphologie flexionnelle riche mais surtout complexe.

Les dictionnaires morphologiques sont utiles pour l'analyse automatique des textes. Nous avons présenté dans cet article la ressource Morfetik, avec son système de traitement qui associe un moteur de flexion, un dictionnaire des formes fléchies, des interfaces de consultation et d'interrogation, ainsi qu'un ensemble d'outils permettant la maintenance et l'exploitation des ressources.

Cependant, pour être utilisé dans un système de traitement automatique des langues, un dictionnaire de formes n'est jamais suffisant : un correcteur orthographique, un générateur de formes liées notamment à des matrices d'affixation permettant de rendre compte des dérivés sont ainsi inévitables. En effet, de nombreuses études ont montré que l'intégration de la morphologie dérivationnelle à la morphologie flexionnelle améliorerait les systèmes de traitement automatique des langues (voir notamment les travaux de JACQUEMIN *et al.*, 1997 ; LEE, 2004). De ce point de vue, une étude supplémentaire doit être menée afin d'utiliser les matrices morphologiques dont dispose Morfetik.

Enfin, ce type de ressources, comme toutes les ressources linguistiques, demande une confrontation continue avec des corpus très divers (GREZKA, CARTIER et MATHIEU-COLAS, 2015). Du point de vue des formes linguistiques, cela

revient à mettre la ressource linguistique en regard avec un corpus continu pour suivre l'évolution fréquentielle des lexies (néologismes, fréquence nulle sur une période, etc.). Une ressource morphologique comme Morfetik est donc l'un des éléments d'un système plus large impliquant un corpus continu et un module de néologismes, ainsi que différents outils pour suivre la fréquence d'usage des lexies du dictionnaire.

BIBLIOGRAPHIE

- CATACH Laurent (2000), Les dictionnaires électroniques des éditions Le Robert, in : DENISE Michel (éd.), *Les dictionnaires électroniques*, Paris, Didier-CIEP, p. 31-42.
- COUGNON Louise-Amélie ; FAIRON Cédric (2009), La mise à jour d'un dictionnaire électronique : Une expérience pédagogique liée à la mise à jour du Delaf, *Arena Romanistica*, 28th Conference on Lexis and Grammar, Bergen (29/09/2009-03/10/2009), vol. 1, no. 4, p. 58-71.
- COURTOIS Blandine (1990), Un système de dictionnaires électroniques pour les mots simples du français, in : COURTOIS Blandine, SILBERZTEIN Max (éd.), *Dictionnaires électroniques du français*, Langue française 87, Paris, Larousse, p. 11-22.
- COURTOIS Blandine ; SILBERZTEIN Max (1990), Dictionnaires électroniques du français, *Langue française* 87, Paris, Larousse, p. 3-4.
- GREZKA Aude (2006), *Les prédicats de perception. Traitement de la polysémie. (Les sens des sens)*. Thèse de doctorat en Sciences du Langage, Université Paris 13, 24 novembre 2006, 834 pages (vol.1 et 2).
- GREZKA Aude ; CARTIER Emmanuel, MATHIEU-COLAS Michel (2015), Dictionnaires morphologiques du français contemporain : présentation de Morfetik, éléments d'un modèle pour le TAL, *TALN 2015*, Caen, 22-25 juin 2015.
- GROSS Gaston (2005), Un dictionnaire électronique des adjectifs du français, *Cahiers de Lexicologie*, Paris, Centre National de la Recherche Scientifique, p. 11-33.
- JACQUEMIN Christian ; KLAUVANS Judith ; TZOUKERMANN Evelyne (1997), Expansion of multi-word terms for indexing and retrieval using morphology and syntax, in : *Proceedings of the 35th Annual Meeting of the Association for Computational*.
- LEE Young-Suk (2004), Morphological Analysis for Statistical Machine Translation, in : Dumais, S. et S. Roukos (éds), *Proceedings of HLT- NAACL 2004*, 57-60.
- MATHIEU-COLAS Michel (2009), Morfetik, une ressource lexicale pour le TAL, *Cahiers de Lexicologie*, Paris, Centre National de la Recherche Scientifique, p. 137-146.