

Dictionnaires distributionnels et étiquetage lexical de corpus

Delphine Reymond

Equipe DELIC – Université de Provence
19, avenue Robert Schuman
13621 Aix-en-Provence - France
reymond@up.univ-aix.fr

Résumé – Abstract

Ce papier présente la première partie d'un travail de thèse qui vise à construire un « dictionnaire distributionnel » à partir d'un corpus de référence. Le dictionnaire proposé est basé sur un ensemble de critères différentiels stricts qui constituent des indices exploitables par des machines pour discriminer le sens des mots en contexte. Pour l'instant, le travail a porté sur 50 000 occurrences qui ont été étiquetées de façon manuelle. Ce sous-corpus pourra servir de corpus d'amorçage pour la constitution d'un corpus étiqueté plus grand, qui pourrait servir à différents tests et travaux sur la désambiguïsation automatique.

This paper presents the first part of a Ph. D. aimed at the construction of a “distributional dictionary” from a reference corpus. The dictionary proposed is based on a set of strict differential criteria that can be used as clues for the discrimination of word senses in context. So far, we have worked on 50,000 occurrences that have been manually tagged. This sub-corpus can be used as a bootstrap corpus for the construction of a larger tagged corpus that could be used in various tests and studies on automatic disambiguation.

Mots-clés – Keywords

Corpus, dictionnaire, étiquetage lexical, information distributionnelle
Corpora, dictionaries, lexical tagging, distributional information

1 Introduction

On a vu apparaître au cours de la dernière décennie des corpus étiquetés du point de vue morpho-syntaxique qui se sont avérés extrêmement utiles pour de nombreuses applications du traitement automatique des langues. Il serait intéressant, afin d'améliorer encore les performances de ces applications, de pouvoir disposer d'informations supplémentaires sur ces corpus, et notamment d'un étiquetage lexical donnant le sens de chaque mot. Toutefois,

malgré des recherches continues depuis une quarantaine d'années (voir l'état de l'art de Ide & Véronis, 1998), la désambiguïsation automatique du sens des mots, nécessaire à tout étiquetage d'envergure, reste un thème de recherche plus qu'une technique opérationnelle. Les résultats de l'évaluation Senseval/Romanseval sur le français¹ ont de ce point de vue été extrêmement modestes (Segond, 2000) : on est bien loin dans ce domaine des chiffres supérieurs à 95% d'efficacité couramment publiés pour l'étiquetage morpho-syntaxique.

Selon Véronis (2001), l'une des difficultés majeures de l'étiquetage lexical automatique, qui explique en grande partie la stagnation des performances des systèmes, réside dans l'inadéquation des dictionnaires traditionnels pour cette tâche, et tout particulièrement dans le manque d'information distributionnelle permettant de relier les sens et les contextes en corpus². Ainsi, par exemple, aucun des dictionnaires français que nous avons pu consulter ne mentionne l'affinité d'un mot aussi simple que *barrage* avec deux prépositions différentes selon son sens : *barrage sur* <une rivière, un fleuve, etc.>, *barrage à* <une idée, un projet, etc.>. Une telle information fournit une base objective au découpage des entrées en « sens », et est de la première importance pour la désambiguïsation par des machines.

La difficulté de discrimination entre les sens possibles d'un mot tels qu'ils sont listés dans un dictionnaire a été mise en évidence par une expérience d'étiquetage manuel portant sur près de 4000 occurrences de 60 mots différents dans un corpus. Six annotateurs, étudiants en linguistique, ont étiqueté manuellement la totalité de ces occurrences à l'aide des sens du *Petit Larousse* (Véronis, 1998, 2000a). L'accord inter-annotateur moyen s'est avéré très faible, de l'ordre de 40% une fois soustrait l'effet du hasard. Pour certains mots tels que *correct*, *historique*, *lancement*, *mettre*, le taux d'accord était à peu près égal à celui qui aurait été obtenu si les annotateurs avaient répondu de façon aléatoire.

Notre travail de thèse s'articule autour de la construction d'un dictionnaire « distributionnel », spécialement adapté au problème de la désambiguïsation par des machines. L'information idéalement contenue dans un tel dictionnaire n'a plus pour but premier la définition du « sens » comme dans les dictionnaires traditionnels, mais vise à organiser les vocables en *lexies* possédant des propriétés distributionnelles cohérentes – propriétés qui peuvent constituer autant d'indices utilisables pour la désambiguïsation automatique.

Ce travail s'inspire largement des idées mises en oeuvre dans la construction du *Dictionnaire Explicatif et Combinatoire* (DEC) de Mel'cuk et son équipe (Mel'cuk, Clas & Polguère, 1995). Toutefois à la différence de Mel'cuk *et al.*, qui utilisent le corpus comme outil de vérification a posteriori après une construction lexicale introspective, nous considérons le corpus comme réservoir d'observations à partir duquel les entrées doivent être élaborées, la démarche introspective étant réduite au minimum. L'entrée de dictionnaire, dans notre approche, est donc vue comme un *modèle*, qui doit rendre compte au mieux des observations, et doit être *prédictif* des observations futures (dans d'autres corpus ou un corpus plus large). Le travail d'élaboration des entrées se fait de façon totalement interactive avec l'étiquetage d'un corpus de référence (le plus large possible), selon une stratégie incrémentale : le corpus

¹ <http://www.up.univ-mrs.fr/~veronis/romanseval>

² Les ressources plus orientées « machine » telles que *WordNet* n'échappent pas à cette critique.

est étiqueté (avec l'assistance d'outils informatiques variés) au fur et à mesure de la construction des entrées, et les entrées sont révisées en fonction des nouveaux contextes rencontrés. A la fin de ce processus, on obtient un dictionnaire dont chacune des lexies est liée à un ensemble de contextes dans le corpus. De même, chaque occurrence dans le corpus est étiquetée par le numéro d'une lexie dans le dictionnaire.

Dans le cadre de notre travail de thèse, il s'agit d'élaborer une méthodologie, et non de construire un dictionnaire dans son intégralité, travail qui ne pourrait se concevoir qu'au sein d'une large équipe lexicographique. Nous avons donc sélectionné un corpus de taille manipulable (cinq millions de mots), et nous avons dans un premier temps traité 60 vocables très polysémiques (20 noms, 20 adjectifs et 20 verbes), qui semblent représenter un large éventail des problèmes lexicographiques pour le traitement automatique des langues. Ces 60 vocables apparaissent dans environ 50 000 contextes au total, qui ont tous été étiquetés selon la stratégie incrémentale décrite précédemment, dans le cadre du projet *Syntsem*, financé par l'ELRA/ELDA. Dans la deuxième partie de notre travail de thèse, nous utiliserons ce sous-corpus comme corpus d'amorçage pour étiqueter semi-automatiquement de nouveaux vocables qui apparaissent au voisinage des vocables déjà étiquetés (par exemple *chute* dans le contexte de *barrage* réfère très probablement au sens « chute d'eau »).

2 Corpus de référence

Le corpus de référence que nous utilisons est composé de cinq parties d'environ un million de mots chacune, de genres variés (chaque partie est désignée par un code d'une lettre) :

- A Littérature (serveur ABU).
- J Institutionnel (Commission européenne).
- M Presse (Le Monde).
- O Ouvrages de sciences humaines (Editions CNRS).
- P Périodiques scientifiques (CNRS et Hermès).

Ce corpus a été choisi pour des raisons évidentes de disponibilité, tout en respectant une diversité de genres intéressante et représentative de nombreuses applications de traitement automatique des langues.

Le corpus a été intégralement étiqueté pour la morpho-syntaxe et lemmatisé à l'aide du logiciel *Cordial 7 Analyseur*, développé et distribué par la société Synapse Développement³. Ce logiciel, qui a l'avantage d'être disponible à un coût modique, constitue l'un des meilleurs étiqueteurs morpho-syntaxiques pour le français : des tests effectués dans notre équipe montrent des taux de performance de l'ordre de 98% (Valli & Véronis, 1999). Le logiciel LoX⁴, outil créé au sein de l'équipe DELIC, dont une des principales fonctions est de permettre l'extraction de concordances (Audibert, 2001), nous a permis d'extraire facilement les fragments du corpus contenant les vocables concernés, sous toutes leurs formes fléchies, et en éliminant les ambiguïtés syntaxiques (par exemple, *porte* verbe vs *porte* nom).

³ <http://www.synapse-fr.com>

⁴ LoX est disponible à l'adresse suivante : <http://laurent.audibert.free.fr/lox.htm>

60 sous-corpus ont été automatiquement créés, chacun composé de la totalité des lignes de concordance d'un des vocables à traiter, dans un format multicolonne (voir Annexe 2) :

- code désignant la partie du corpus (A, J, M, O, P) ;
- numéro de paragraphe de l'occurrence ;
- numéro de caractère dans le paragraphe ;
- contexte gauche (50 mots) ;
- occurrence ;
- colonne vide (destinée à recevoir l'étiquette lexicale) ;
- contexte droit (50 mots).

Les sous-corpus ainsi obtenus sont d'une taille beaucoup plus réduite que le corpus initial de cinq millions de mots, ce qui facilite l'ensemble des manipulations, et leur format tabulaire est parfaitement adapté à la stratégie d'étiquetage « vertical » que nous avons adoptée. En effet, nous n'étiquetons pas chaque mot dans l'ordre linéaire du texte, mais par vocable : toutes les cooccurrences d'un vocable sont traitées simultanément, à l'aide d'une présentation sous forme de concordance (Annexe 2). Cette stratégie permet une meilleure cohérence des décisions d'étiquetage sur l'ensemble du corpus, et facilite grandement la tâche de l'annotateur, qui peut utiliser des fonctions de tri et de sélection diverse et étiqueter des paquets entiers d'occurrences qui ont des contextes gauche et/ou droit analogues.

3 Vocables traités

Les 60 vocables que nous avons traités sont ceux qui ont été utilisés dans l'expérience d'étiquetage manuel mentionnée précédemment (Véronis, 1998, 2000a). Ces mêmes vocables ont également été utilisés dans l'évaluation de l'étiquetage lexical Senseval/Romanseval (Segond, 2000), ainsi que dans l'évaluation de l'alignement de textes parallèles français-anglais Arcade (Véronis, 2000b). Notre corpus permettra donc des comparaisons intéressantes avec d'autres travaux. En particulier, on pourra vérifier si notre « dictionnaire distributionnel » permet une amélioration de l'étiquetage, tant manuel qu'automatique.

Ces 60 vocables se répartissent en 20 noms, 20 adjectifs et 20 verbes, et ont été choisis d'une part sur la base de leur fréquence assez élevée, et d'autre part de leur forte polysémie (figure 1).

Noms	<i>barrage, chef, communication, compagnie, concentration, constitution, degré, détention, économie formation, lancement, observation, organe, passage, pied, restauration, solution, station, suspension, vol</i>
Verbes	<i>arrêter, comprendre, conclure, conduire, connaître, couvrir, entrer, exercer, importer, mettre, ouvrir, parvenir, passer, porter, poursuivre, présenter, rendre, répondre, tirer, venir</i>
Adjectifs	<i>biologique, clair, correct, courant, exceptionnel, frais, haut, historique, plein, populaire, régulier, sain, secondaire, sensible, simple, strict, sûr, traditionnel, utile, vaste</i>

Figure 1 : liste des 60 vocables traités

4 Critères différentiels

Nous avons donc entrepris la construction des entrées du dictionnaire en utilisant systématiquement des critères distributionnels stricts, tels que (1) les contraintes syntaxiques observées pour un sens donné (par exemple, type de préposition, comme dans *barrage sur*, *barrage à*) ; (2) les restrictions de sélection (par exemple on *construit* un barrage sur un fleuve, on *dresse* un barrage sur une route, on *fait barrage* à l'opposition) ; (3) des tests de commutation de synonymes stricts (on peut remplacer « faire barrage à l'opposition » par « faire *obstacle* à l'opposition », mais on ne peut avoir de façon naturelle « construire un *obstacle* sur un fleuve »), etc. Ces critères sont des *critères différentiels*, en ce sens qu'ils permettent d'opposer les diverses classes d'emplois du vocable, et donc de le découper en lexies cohérentes. Nous ne pouvons dans le cadre de ce papier détailler la totalité des critères utilisés, et nous nous bornerons à un certain nombre d'exemples significatifs :

- **valence (type et nombre des compléments)**

comprendre que..., *comprendre qqun*
barrage sur qqch, *barrage à qqch*

- **traits**

- \pm comptable

la communication (d'entreprise), *la/les communications (téléphoniques)*

- \pm sexué

le chien/la chienne (qui aboie), *le chien/*la chienne (de fusil)*

- \pm massif

du vin, *un (grand) vin*

- \pm animé

arrêter une date, *arrêter quelqu'un*

- etc.

- **auxiliaire être / avoir**

Je suis passé sous l'échelle, *j'ai passé le baccalauréat*

- **emploi réfléchi**

conduire/se conduire (bien), *porter/se porter (candidat)*

- **dérivation active (peut être utilisée de façon productive pour reformuler l'énoncé)**

importer : *l'importation de / avoir de l'importance*
élever : *élevage de cochons*, **élevage d'enfants*

- **hyperonymes**

*barrage sur un fleuve : ouvrage, barrage sur une route : *ouvrage*

- **synonymes**

*faire barrage/obstacle à l'opposition, construire un barrage/*obstacle sur un fleuve*

- **antonymes, converses**

*boissons fraîches/chaudes, roses fraîches/fanées
importer (des marchandises)/exporter, importer (être important)/*exporter*

- **restrictions de sélection, classes d'objets**

*construire un barrage (sur un fleuve), dresser un barrage (sur une route),
faire barrage à (un projet, des idées)
lancement de <fusée, missile...>, lancement de <événement, projet...>
lire <expression, sentiment...> (sur un visage), lire <livre, lettre,
journal...>*

- **figements, semi-figements**

rendre l'âme, chef d'œuvre, haut en couleurs

- **etc.**

L'information ainsi dégagée est incorporée dans chaque entrée lexicale. Nous présentons en Annexe 1 une entrée (simplifiée), concernant le vocable **IMPORTER** (on trouvera en Annexe 2 l'étiquetage d'un fragment du corpus correspondant à ce vocable).

5 Conclusion

Nous avons exposé dans ce papier la première partie d'un travail de thèse qui vise à construire de façon incrémentale un dictionnaire « distributionnel » à partir d'un corpus de référence. Ce travail part d'un constat d'inadéquation des dictionnaires classiques (et de ressources informatisées telles que *WordNet*) pour la désambiguïsation du sens des mots par des machines et l'étiquetage lexical automatique de corpus. Le dictionnaire proposé est basé sur un ensemble de critères différentiels stricts, qui permettent d'opposer les différentes classes d'emploi de chaque vocable (ou lexies). Ces critères, codés dans l'entrée lexicale, peuvent constituer ultérieurement autant d'indices exploitables par des machines pour discriminer le sens des mots en contexte. Pour l'instant, le travail a porté sur 60 vocables très polysémiques, représentant 50 000 occurrences dans un corpus de cinq millions de mots, qui ont été étiquetées de façon manuelle. Nous espérons que ce sous-corpus pourra servir de corpus d'amorçage pour l'étiquetage semi-automatique des mots du voisinage de ceux déjà étiquetés, permettant, de façon relativement économique, de constituer un corpus étiqueté plus grand, qui pourrait servir à différents tests et travaux sur la désambiguïsation automatique.

Références

Audibert, L. (2001, 2-5 juillet). LoX : outil polyvalent pour l'exploration de corpus annotés. *Actes de RECITAL'2001*, Tours.[Présent volume].

Ide, N., & Véronis, J. (1998). Introduction to the special issue on word sense disambiguation: the state of the art. *Computational Linguistics*, 24(1), 1-40. <http://www.up.univ-mrs.fr/~veronis/pdf/1998wsd.pdf>

Mel'cuk, I., Clas, A., & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Editions Duculot (Coll. Universités Francophones).

Segond, F. (2000). Framework and results for French. *Computers and the Humanities*, 34(1/2), 49-60 [special issue on Senseval].

Valli, A., & Véronis, J. (1999). Etiquetage grammatical de corpus oraux: problèmes et perspectives. *Revue Française de Linguistique Appliquée*, IV(2), 113-133. <http://www.up.univ-mrs.fr/~veronis/pdf/1999rfla.pdf>

Véronis, J. (1998). A study of polysemy judgements and inter-annotator agreement, *Programme and advanced papers of the Senseval workshop* (pp. 2-4). Herstmonceux Castle (England). <http://www.up.univ-mrs.fr/~veronis/pdf/1998senseval.pdf>

Véronis, J. (2000a). Sense tagging: Don't look for the meaning but for the use, *Computational Lexicography and Multimedia Dictionaries (COMLEX'2000)* (pp. 1-9). Kato Achia (Greece). <http://www.up.univ-mrs.fr/~veronis/pdf/2000comlex.pdf>

Véronis, J. (2000b). Evaluation of parallel text alignment systems: the ARCADE project. In J. Véronis (Ed.), *Parallel text processing: Alignment and use of translation corpora* (pp. 369-388). Dordrecht: Kluwer Academic Publishers.

Véronis, J. (2001). Sense tagging: does it make sense? *Corpus Linguistics'2001* (pp. in press). Lancaster, U.K. <http://www.up.univ-mrs.fr/~veronis/pdf/2001-lancaster-sense.pdf>

Annexe 1 - Exemple d'entrée (simplifiée)

IMPORTER, v.

1. IMPORTER (sans objet)

1.1 Qqch importe [à qqun, pour qqun]

<i>TNom :</i> x IMPORTE à/pour y \Rightarrow x a de l'IMPORTANCE pour y
<i>TAdj :</i> x IMPORTE à/pour y \Rightarrow x a est IMPORTANT pour y

Ex : le contenu importe plus que le contenant
la forme du gouvernement importe peu
Les actions qui importent pour eux
ce qui m'importe, c'est...

1.2 Qu'/peu importe qqch/que P [à qqun, pour qqun]

Ex : Peu importait le procédé à Bouvard
Qu'importe la croyance
Que m'importe la tempête, si j'ai la boussole

1.3 il importe [à qqun, pour qqun] de Vinf /que P (impersonnel)

Ex : Ils savent, dès avant d'ouvrir un livre, ce qu'il importe d'en penser
il nous importe donc de montrer

1.4 N'/peu importe Pro

Ex : N'importe lequel, laquelle, quel(s), quelle(s), qui, quoi...
N'importe où, quand, comment...

1.5. N'/peu importe !

2. IMPORTER QQCH [QQPART]

<i>TNom :</i> x IMPORTE y \Rightarrow l'IMPORTATION de y par x
<i>Conv :</i> EXPORTER

2.1. Importer <objets, marchandises...> [dans/vers <pays, région...>]

Ex : Moscou a commencé à importer ces produits...
des institutions qui importent des armes

2.2. Importer <population, main d'œuvre...> [dans/vers <pays, région...>]

Ex : on ne peut pas importer toute la population

2.3. Importer <concepts, idées...> [dans/vers <culture, discipline...>]

Ex : il importe un vocabulaire et une méthode propres aux sciences naturelles vers les sciences sociales.
Georges Frèche décide d'importer acteurs et structures culturelles qui font défaut à la cité

Annexe 2 - Exemple d'étiquetage

Fichier	P	POS	Contexte_gauche	Mot	Etiquette	Contexte_droit
M	6620	41	. C ' est l ' avenir qui	importe	1.1	, et cet avenir doit leur être montré tel qu ' il est :
O	11509	45	. Ce qui	importe	1.1	est le cheminement décrit , les obstacles rencontrés , les hésitations , les
P	9233	180	. Ce qui	importe	1.1	, c ' est de comprendre comment tous nos discours , même les plus quoti
P	5532	178	. Ce qui	importe	1.1	ce n ' est pas l ' adéquation des énoncés politiques avec la réalité qu ' elle
A	12580	58	. Ce qui	importe	1.1	, c ' est le chef .
M	6339	51	. Le postulat de M . Pons est que ce fossé	importe	1.1	peu pourvu que l ' Etat manifeste assez d ' autorité pour imposer la politici
O	7541	149	à réalisé ou qu ' il se réalise dans le futur	importe	1.1	peu , pourvu que sa survenue soit certaine .
A	9067	20	. « La forme du gouvernement	importe	1.1	peu !
A	11684	20	. Le fond de l ' histoire	importe	1.1	peu ;
A	8452	6	. - - « Qu '	importe	1.2	!
A	9490	34	. Que ce soit une apparence , qu '	importe	1.2	! il s ' agit de la produire .
A	9660	8	. Elle ou un autre , qu '	importait	1.2	! le principal étant de sortir d ' embarras .
O	1051	119	. Peu	importe	1.2	, à cet égard , que le rapport ne soit pas encore divulgué .
O	2126	34	. Peu	importe	1.2	, après cela qu ' infidèles à leurs méthodes , qu ' infidèles à leurs doctrine
A	1077	89	; peu	importe	1.2	, après tout , puisqu ' il a dû livrer enfin le secret de ses sources !
M	1773	22	it sur la scène , mais la société , bon , peu	importe	1.2	, ce n ' est pas ici un discours de François Mitterrand .
A	4268	4	: " Qu '	importe	1.2	, d ' ailleurs , un peu plus ou un peu moins de génie , puisque tout doit fin
A	17039	69	ous - même , agacé peut - être , mais peu	importe	1.2	, et , par ce travail , vous accroissez , bon gré , mal gré , la vie de Poil de
P	4900	159	. Mais qu '	importe	1.2	, il faut que ce soit dans le domaine public .
M	10081	42	. Il lui	importait	1.2	, malgré tout , de savoir encore si Mme Lea Katz avait remarqué , à l ' UC
M	701	78	. Qu '	importe	1.2	, répondent les dirigeants de la Cinq , les techniciens italiens de M . Silvid
P	7283	257	. Mais n '	importe	1.3	- t - il pas de reconnaître également qu ' un chercheur met de l ' avant de
O	2758	88	. Aussi	importe	1.3	- t - il que ne soient pas gaspillés les moyens et les efforts consacrés à l '
P	3475	234	. De même , il	importe	1.3	" d ' expliquer " le journal » .
O	1206	70	. Il	importe	1.3	, à cet égard , de relever que le CPI , entres autres considérations , prend
O	382	18	ction des connaissances étant effectué , il	importe	1.3	, à présent , d ' examiner la question de la protection de la forme de l ' œu
A	4882	111	oit descendu dans sa tombe , c ' est qu ' il	importe	1.3	, après ce que vous m ' avez dit , que vous sachiez bien qui je suis , afin
A	2227	43	. Ce qu ' il	importe	1.3	, c ' est de gagner du temps .
A	25247	43	. " Il	importe	1.3	, dit M . H . Poincaré (*) , de ne pas multiplier les hypothèses outre mes
M	2490	59	rêt assassins étrangent l ' activité et qu ' il	importe	1.3	avant tout de les faire baisser ?
O	2794	113	? Il	importe	1.3	bien de savoir si nous faisons de la sociologie .
P	8863	2	. Il	importe	1.3	cependant de ne pas assimiler le souci de la mise en relation de la sensiti
P	7248	231	ues psycho - ou socio - linguistiques qu ' il	importe	1.3	d ' analyser dans leur fonctionnement effectif ;

Fichier	P	POS	Contexte_gauche	Mot	Etiquette	Contexte_droit
P	3891	12	r ailleurs , dans le cas qui nous occupe , il	importe	1.3	d ' apporter une distinction analytique supplémentaire pour tenir compte d
P	8237	52	. Mais il	importe	1.3	d ' introduire une distinction .
M	5044	16	ut bien " être une vedette . . . mais pas n '	importe	1.4	comment " et surtout il ne veut pas " en avoir bavé pour que d ' autres en
A	18223	5	? - - N '	importe	1.4	comment , par tout le monde , par le premier venu , par moi , tiens !
A	5741	22	je vous rencontrerai . . . dans la rue . . . n '	importe	1.4	où . . . à l ' heure que vous voudrez . . . pourvu que je vous voie .
P	4771	179	. On ne porte pas une robe n '	importe	1.4	où et n ' importe comment .
P	3788	252	lement vendus pas le monde associatif (ba	importe	1.4	quel catalogue de vente par correspondance :
P	5680	214	Un enfant pourrait en faire autant » , « N '	importe	1.4	qui pourrait en faire autant » , qui émaillent les livres d ' or de tant d ' expo
M	10562	97	l ' ex - CERES n ' est pas prêt à signer n '	importe	1.4	quoi) .
A	7637	35	. La terre buvait l ' eau , n '	importe	1.5	!
A	7435	27	; n '	importe	1.5	!
A	10084	3	. » N '	importe	1.5	!
A	8069	3	. N '	importe	1.5	! cette fin du monde , si lointaine qu ' elle fût , les assombrit - - et côte à c
A	7915	6	. - - « N '	importe	1.5	! il en a besoin !
A	8315	3	. N '	importe	1.5	! ils devaient le rendre .
O	9677	23	igeaient un charbon de qualité qu ' il fallait	importer	2.1	, alors que les centrales EDF acceptent tous les charbons) , cette affirma
O	8754	73	. En 1913 , il fallait déjà	importer	2.1	5 000 000 tonnes de première qualité (criblés 3 / 4 gras et tout - venant q
O	8721	89	ds sont consommés alors que , dit - il , on	importe	2.1	6 milliards de francs de charbon pour la traction à vapeur .
O	8754	12	de charbon de première qualité qu ' il faut	importer	2.1	à prix fort , le chemin de fer français a besoin de plus en plus d ' énergie
O	8730	27	t toujours lié au prix du charbon qu ' il faut	importer	2.1	car les locomotives à vapeur des réseaux français exigent un charbon de
O	3361	65	alement est moins coûteux que ce qui est	importé	2.1	d ' Europe .
O	3343	45	d ' Argentine qui souhaitait ne plus avoir à	importer	2.1	de pare - chocs en plastique .
O	3073	121	cialistes de n ' importe où , on ne peut pas	importer	2.2	toute la population .
P	8007	111	e , maire socialiste élu en 1977 , décide d '	importer	2.3	acteurs et structures culturels qui font défaut à la cité .
M	3511	90	. C ' est l ' américain Dell qui a	importé	2.3	cette méthode en France .
O	1976	16	méthodologie aux sciences naturelles et l '	importent	2.3	dans le domaine de la critique littéraire ou artistique voire dans leur techn
P	4787	95	. Il est , certes , impossible d '	importer	2.3	directement la robe décontractée et décolletée mais il est possible , une f
A	1214	14	- - Une fameuse mode , dit Joe , et que j '	importerai	2.3	en Angleterre ;
O	2279	222	la méthode et le vocabulaire de Darwin et	importer	2.3	l ' idée d ' évolution dans les sciences économiques et historiques .
O	2191	25	s organicistes , une sociobiologie , mais il	importe	2.3	un vocabulaire et une méthode propres aux sciences naturelles vers les s

Figure 2. Exemple d'étiquetage « vertical »