

Distribution des Systèmes Lexicaux

Version 3.1

Sandrine Ollinger

sandrine.ollinger@atilf.fr

Alain Polguère

alain.polguere@univ-lorraine.fr

26 avril 2024

Ressource distribuée sous licence :
Creative Commons – Attribution 4.0 International (CC BY 4.0)



© 2024 Sandrine Ollinger & Alain Polguère

Table des matières

Introduction	2
1 Historique des versions	2
1.1 Version 3.1 : changements par rapport à la version 3	2
1.2 Version 3.0 : changements par rapport à la version 2.1	2
1.3 Version 2.1 : changements par rapport à la version 2.0	3
1.4 Version 2.0 : changements par rapport à la version 1.0	3
2 Organisation générale des fichiers d’export	4
3 Nœuds lexicaux du graphe	4
4 « Entrées » du Système Lexical	5
5 Structure polysémique des vocables	6
6 Caractéristiques grammaticales	8
7 Signifiants lexicaux	11
8 Caractérisation sémantique	13
8.1 Étiquettes sémantiques	13
8.2 Formes propositionnelles	17
8.3 Définitions	18
9 Liens de fonctions lexicales (FL)	19
10 Exemples lexicographiques	23
Références	25

Introduction

Cette documentation décrit le contenu de la version 3.1 des fichiers de distribution sur la plateforme Ortolang des *Systèmes Lexicaux*¹, les réseaux lexicaux développés au laboratoire ATILF. Pour l’instant, seules sont distribuées les données du *Réseau Lexical du Français*, désormais *RL-fr*. Ce sont celles utilisées dans les illustrations du présent document.

La lecture de ce document et l’exploitation des données lexicales qu’il présente présuppose une familiarisation minimale avec l’approche des Systèmes Lexicaux. Pour cela, nous recommandons d’effectuer, dans l’ordre, les lectures ci-dessous.

1. Lexicographie des Systèmes Lexicaux : Polguère (2014).
2. Projet RELIEF, dans le cadre duquel a débuté le programme de construction des Systèmes Lexicaux : Lux-Pogodalla et Polguère (2011).
3. Éditeur lexicographique Dicet, utilisé pour construire les bases lexicales SQL dont sont extraites les données présentées ici : Gader et al. (2012).
4. Structure de graphe petit-monde des Systèmes Lexicaux : Gader et al. (2014b).

Pour tout ce qui relève de la terminologie linguistique employée dans le présent document, on peut se reporter à Polguère (2016) – *cf. Index des notions*, pp. 371–379 de l’ouvrage.

Conventions d’écriture. Les notions linguistiques importantes apparaissent *en italique sans empattement* la première fois qu’elles sont utilisées. De façon plus générale, nous adoptons ici les conventions d’écriture de Polguère (2016).

Vérification formelle des données. Cette documentation énumère les vérifications de la validité formelle des données effectuées à chaque mise à jour dans des sous-sections intitulées *Vérifications effectuées à chaque mise à jour*. Les vérifications introduites par le symbole \odot donnent lieu à une correction systématique avant distribution de la mise à jour.

1 Historique des versions

1.1 Version 3.1 : changements par rapport à la version 3

La version 3.1 comporte uniquement un enrichissement des données linguistiques de la version 3.0, sans changement dans le nombre de fichiers, la structure des données, etc.

1.2 Version 3.0 : changements par rapport à la version 2.1

- La version 3.0 comporte un enrichissement des données linguistiques de la version 2.1.
- Elle comporte des changements dans l’organisation et le contenu du modèle des caractéristiques grammaticales (*cf.* section 6).
- Certains changements ont été apportés au modèle des fonctions lexicales (*cf.* section 9) pour tenir compte de la version du modèle présentée dans l’article de référence Mel’čuk et Polguère (2021).

¹<https://hdl.handle.net/11403/lexical-system-fr>

1.3 Version 2.1 : changements par rapport à la version 2.0

La version 2.1 comporte uniquement un enrichissement des données linguistiques de la version 2.0, sans changement dans le nombre de fichiers, la structure des données, etc.

1.4 Version 2.0 : changements par rapport à la version 1.0

Note. La numérotation des fichiers a changé, suite à l'ajout de deux fichiers de données.

- Il n'y a plus de DTD incluses dans les fichiers XML, mais un seul fichier relax NG compact pour tous les fichiers XML.
- Le typage des données présenté dans la documentation, qui était exprimé en types SQL pour la version 1.0, est désormais exprimé en types de données XML, selon les recommandations de la W3C².
- Les ID sont désormais tous préfixés par de l'information sur la nature de l'entité qu'ils identifient. Ils possèdent la structure suivante :

`<Type de ressource>:<Langue>:<Type d'entité>:<No. ID>.`

Ils ne sont donc plus des entiers, mais des chaînes de caractères. Par exemple, le nœud lexical 26162 issu du RL-fr a pour ID `ls:fr:node:26162`, où `ls` indique que la ressource est un Système Lexical (*Lexical System*), `fr` qu'il s'agit du RL-fr et `node` que l'entité correspondante est un nœud lexical du graphe.

- Le fichier `01-nodes.csv` a été augmenté de trois champs :
 1. date de la dernière mise à jour ;
 2. heure de la dernière mise à jour ;
 3. nom lexicographique complet du nœud lexical dans un format HTML.
- Le fichier `06-lsgramcharac-rel.csv` a connu quatre changements significatifs :
 1. ajout des variables éventuelles dans les noms de marques d'usage ;
 2. ajout des variables éventuelles dans les autres caractéristiques grammaticales ;
 3. correction orthographique du nom de champ `embededlex` → `embeddedlex` ;
 4. ajout des formes des lexies dans le champ `embeddedlex` ;
 5. réorganisation et nettoyage de la hiérarchie des caractéristiques grammaticales.
- Dans le fichier `11-lspropform-rel.csv`, nettoyage des données des champs `actantslist` et `tildevalue` (correction de la gestion des variables scindées et normalisation des espaces).
- Pour la distribution des définitions lexicographiques, ajout des deux fichiers de données `12-definition.dtd` et `13-ls-def.csv`.
- Dans le fichier `14-lslf-model.xml`, nouveau format HTML pour les noms de fonctions lexicales.
- Dans le fichier `18-lsex-rel.csv`, le mode de spécification de la position des occurrences dans les exemples a changé.

²<https://www.w3.org/TR/xmlschema11-2/>

Bien entendu, cette version, contient de nouvelles données lexicographiques et de nombreuses améliorations ont été apportées dans les descriptions déjà existantes. Remerciements pour la version 2.0. :

Merci Yannick Chudy, merci Michaëlle Dubé, merci Guy Lapalme, merci François Lareau, merci Sylvain Pogodalla, merci, merci, merci !

2 Organisation générale des fichiers d'export

La version courante de l'export³ comporte le fichier `.rnc`, les 11 fichiers `.csv`, les 6 fichiers `.xml` et le fichier `.dtd` ci-dessous, tous encodés en UTF-8.

- **Schéma de données**
00-lsschema.rnc
- **Nœuds lexicaux du graphe**
01-lsnodes.csv
- **Entrées du Système Lexical**
02-lsentries.csv
- **Structure polysémique des vocables**
03-lscopolysemy-model.xml
04-lscopolysemy-rel.csv
- **Caractéristiques grammaticales**
05-lsgramcharac-model.xml
06-lsgramcharac-rel.csv
- **Signifiants lexicaux**
07-lswordform-model.xml
08-lswordforms.csv
- **Caractérisation sémantique**
Étiquettes sémantiques
09-lssemlabel-model.xml
10-lssemlabel-rel.csv
Formes propositionnelles
11-lspropform-rel.csv
Définitions
12-definition.dtd
13-lsdef.csv
- **Liens de fonctions lexicales**
14-lslf-model.xml
15-lslf-rel.csv
- **Exemples lexicographiques**
16-lsexsource-model.xml
17-lsex.csv
18-lsex-rel.csv

Les fichiers `.xml` contiennent le **modèle lexicologique** (hiérarchie des caractéristiques grammaticales, système des fonctions lexicales, etc.). La structure de ces fichiers est précisée sous forme de schéma *Relax NG compact* dans le fichier `00-lsschema.rnc`.

Les fichiers `.csv` contiennent les données du **modèle lexicographique** lui-même, qui décrivent le lexique de la langue (les entités lexicales, leurs propriétés individuelles, les liens qui les unissent, etc.). Leurs champs sont séparés par des tabulations et entourés de guillemets dactylographiques doubles ("`...`").

Note. L'utilisation des guillemets dactylographiques doubles et des apostrophes droites par les lexicographes est prohibée. Une attention particulière est portée au remplacement de ces signes par les guillemets français et l'apostrophe courbe dans les données du RL-fr.

Le format d'export des Systèmes Lexicaux est le même, quelle que soit la langue modélisée.

3 Nœuds lexicaux du graphe

Ce segment de l'export contient la définition de l'ensemble des nœuds du graphe. Ce sont par défaut des *lexies*, = *unités lexicales*, de la langue. Il peut cependant aussi s'agir de *clichés linguistiques*, de *collocations* ou même, marginalement, d'*expressions libres* (Mel'čuk, 2012). Cette information est encodée dans un unique fichier : `01-lsnodes.csv`.

³L'export est réalisé en PHP à l'aide de trois scripts programmés par S. Ollinger : `mainExportSpiderlex.php`, `DBAccessExport.php` et `GenerateExport.php`.

Fichier 01-lsnodes.csv

id	entry	lexnum	status	%	update_date	update_time	lexname
ls:fr:node:26162	ls:fr:entry:26164	l.1	2	100	2021-03-13	09:45:08	à1.1
ls:fr:node:26163	ls:fr:entry:26165	1	2	100	2017-03-07	08:28:24	à propos1
ls:fr:node:26164	ls:fr:entry:26166	ll.2	2	100	2017-12-03	21:05:52	abandonnerll.2

FIG. 1 : Extrait du fichier 01-lsnodes.csv

id	token	ID du nœud lexical (typiquement, une lexie).
entry	token	ID de l'entrée du Système Lexical correspondante (typiquement, un vocable – voir section 4 plus bas).
lexnum	string	Numéro lexicographique du nœud lexical.
status	integer	Statut d'avancement du travail lexicographique.
%	integer	Degré de confiance de l'information.
update_date	date	Date de la dernière mise à jour.
update_time	time	Heure de la dernière mise à jour.
lexname	token	Nom lexicographique du nœud lexical, en format HTML.

TAB. 1 : Description des données du fichier 01-lsnodes.csv

Le paramètre `status` prend une des quatre valeurs numériques ci-dessous :

- 3 → présence d'informations lexicographiques minimales ;
- 2 → description lexicographique véritablement débutée ;
- 1 → description lexicographique proche de la finalisation ;
- 0 → description lexicographique finalisée.

Vérifications effectuées à chaque mise à jour

- ⊙ Y a-t-il des lexies dont le numéro lexicographique (`lexnum`) est 0 ? Il s'agit de nœuds du graphe signalés par les lexicographes comme devant être détruits par la personne en charge de l'administration du modèle.

4 « Entrées » du Système Lexical

Un unique fichier, `02-lsentries.csv`, contient la spécification des « entrées » du Système Lexical, qui sont, dans la majorité des cas, des *vocables* de la langue (comme dans un dictionnaire courant). Chaque nœud du graphe étant associé à une entrée, il existe bien entendu marginalement des entrées correspondant à des collocations, des syntagmes libres, etc., et non à des vocables.

Dans ce fichier, sont notamment stockés les *noms lexicographiques* des vocables. Ainsi, la Figure 2 ci-dessous montre comment sont spécifiés les noms lexicographiques des vocables À, À PROPOS, ADULTE_{Adj} et S'ALLONGER¹ (qui possède un vocable homonyme S'ALLONGER²).

Fichier 02-lsentries.csv

id	addtoname	name	subscript	superscript	status	%
ls:fr:entry:26164		à			1	100
ls:fr:entry:26165		à propos			2	100
ls:fr:entry:26215		adulte	Adj		2	100
ls:fr:entry:26273	s'	allonger		1	2	100

FIG. 2 : Extrait du fichier 02-lsentries.csv

id	token	ID de l'entrée (typiquement, un vocable)
addtoname	string	Chaîne de caractères qui débute le nom lexicographique de l'entrée et qui ne doit pas entrer en ligne de compte dans le classement alphabétique de la nomenclature du Système Lexical : voir <i>se</i> et <i>s'</i> , pour les verbes pronominaux français.
name	string	Corps du nom lexicographique de l'entrée.
subscript	token	Indice qui s'ajoute au nom lexicographique de l'entrée pour le spécifier grammaticalement, par exemple pour distinguer un vocable adjectival d'un vocable nominal nommé par le même signifiant linguistique (PETIT _(Adj) ~ PETIT _(N)).
superscript	token	Exposant (normalement, numérique) qui s'ajoute au nom lexicographique de l'entrée pour distinguer des homonymes (AVOCAT ¹ ~ AVOCAT ²).
status	integer	Statut d'avancement du travail lexicographique. Prend les mêmes valeurs que le paramètre status de la Figure 1, décrites plus haut p. 5.
%	integer	Degré de confiance de l'information.

TAB. 2 : Description des données du fichier 02-lsentries.csv

5 Structure polysémique des vocables

La *structure polysémique* des vocables est modélisée dans deux fichiers :

- 03-lscopolysemy-model.xml contient la définition des différents types et sous-types de liens de *copolysemy* ;
- 04-lscopolysemy-rel.csv contient l'ensemble des liens de copolysemy identifiés dans le Système Lexical.

L'ensemble des liens de copolysemy unissant les acceptions d'un vocable donné constitue la structure polysémique de ce vocable. Notons que cette structure est hiérarchique, mais non nécessairement arborescente : il peut exister des cas (rares) de lexies qui sont subordonnées à plus d'une autre acception de leur vocable (Polguère, 2018).

Fichier 03-lscopolysemy-model.xml

```

<model>
  <type id="ls:fr:cptype:3" name="Causation" order="1" semantics="2" derivation="1"/>
  <type id="ls:fr:cptype:6" name="Résultat" order="2" semantics="2" derivation="1"/>
  <type id="ls:fr:cptype:7" name="Conversion" order="3" semantics="2" derivation="1"/>
  <type id="ls:fr:cptype:4" name="Spécialisation" order="4" semantics="2" derivation="1">
    <subtype id="ls:fr:cpstype:10" name="Sous-sens"/>
  </type>
  ...
</model>

```

Figure 3: Extrait du fichier 03-lscopolysemy-model.xml

id	token	ID du lien de copolysémie.
name	string	Nom du lien de copolysémie.
order	integer	Position absolue du lien de copolysémie dans une énumération des liens.
semantics	integer	Poids sémantique du lien de copolysémie : 0 = poids nul, 1 = poids intermédiaire et 2 = poids maximal.
derivation	boolean	Pour un lien de copolysémie $L_1 \rightarrow L_2$, l'acceptation L_2 est/n'est pas considérée comme construite sémantiquement à partir de l'acceptation L_1 .

TAB. 3 : Description des données du fichier 03-lscopolysemy-model.xml

Fichier 04-lscopolysemy-rel.csv

source	target	type	subtype
ls:fr:node:44006	ls:fr:node:44007	ls:fr:cptype:1	ls:fr:cpstype:4
ls:fr:node:32937	ls:fr:node:28866	ls:fr:cptype:5	
ls:fr:node:26291	ls:fr:node:30859	ls:fr:cptype:1	ls:fr:cpstype:7

FIG. 4 : Extrait du fichier 04-lscopolysemy-rel.csv

source	token	ID de la source du lien de copolysémie.
target	token	ID de la cible du lien de copolysémie.
type	token	ID du type du lien de copolysémie, tel que défini par <code>type</code> dans le modèle (voir plus haut, fichier 03-lscopolysemy-model.xml).
subtype	token	ID du sous-type du lien de copolysémie, tel que défini par <code>subtype</code> dans le modèle.

TAB. 4 : Description des données du fichier 04-lscopolysemy-rel.csv

Le lien logique entre `type` et `subtype` n'est pour l'instant pas encodé de façon explicite dans la base lexicographique. Il n'est donc pas récupérable automatiquement⁴. Comme l'éditeur

⁴Un bricolage jouant entre `position` et `id` se fait actuellement dans le script d'export afin d'associer le bon `subtype` au `type` correspondant dans la structure XML du fichier 03-lscopolysemy-model.xml.

lexicographique ne dispose pas de l'information nécessaire pour contrôler ce lien logique, il peut arriver que les lexicographes créent des associations <type>:<subtype> invalides telles que Métaphore:Ensemble de⁵.

Vérifications effectuées à chaque mise à jour

- Y a-t-il de mauvaises associations entre types et sous-types de liens de copolysémie ?

6 Caractéristiques grammaticales

Les caractéristiques grammaticales associées à chaque nœud du Système Lexical (marques d'usage, partie du discours, etc.) sont organisées au sein d'une hiérarchie. La caractérisation grammaticale est exportée dans deux fichiers :

- 05-lsgramcharac-model.xml contient la hiérarchie des caractéristiques grammaticales disponibles pour l'étiquetage grammatical des nœuds lexicaux ;
- 06-lsgramcharac-rel.csv contient l'ensemble des étiquetages grammaticaux (associations entre les nœuds lexicaux et leurs caractéristiques grammaticales).

Fichier 05-lsgramcharac-model.xml

```
<?xml version="1.0" encoding="utf-8"?>
<?xml-model href="00-ls-schema.rnc" type="application/relax-ng-compact-syntax"?>
<model>
  <characteristic id="ls:fr:gc:2" name="MARQUES D'USAGE" type="0" status="0">
    <characteristic id="ls:fr:gc:3" name="Marques d'usage langagier" type="0" status="0">
      <characteristic id="ls:fr:gc:6" name="spéc" type="1" status="0"/>
      <characteristic id="ls:fr:gc:272" name="(spéc)" type="1" status="0">
        <![CDATA[<html><body><p>(spéc) = spéc ou gén</p></body></html>]]>
        ...
      </characteristic>
      ...
    </characteristic>
    ...
  </characteristic>
  ...
</model>
```

FIG. 5 : Extrait du fichier 05-lsgramcharac-model.xml

⁵Ensemble de est un sous-type valide pour Métonymie et non pour Métaphore.

id	token	ID de la caractéristique grammaticale.
name	string	Nom de la caractéristique grammaticale.
type	integer	Type de la caractéristique grammaticale. Trois types sont utilisés : 0 pour ‘autre caractéristique grammaticale’ (valeur par défaut), 1 pour ‘marque d’usage’ et 2 pour ‘partie du discours’.
status	integer	Statut lexicographique de la caractéristique grammaticale. Dans l’état courant du modèle, seulement deux valeurs sont utilisées : 0 = caractéristique grammaticale validée <i>vs</i> 1 = caractéristique grammaticale non encore validée. Dans l’export courant, seules sont présentes les caractéristiques grammaticales validées – c’est-à-dire, de statut 0.
CDATA	string	Optionnel : commentaire explicatif sur la caractéristique grammaticale. Ce type de commentaire est particulièrement important pour les caractéristiques qui ne relèvent pas de la terminologie linguistique la plus courante : par exemple, il est pertinent d’avoir un commentaire pour le statut sémantique pragmonyme , mais pas pour le genre grammatical masc .

TAB. 5 : Description des données du fichier `05-lsgramcharac-model.xml`

L’algorithme de création du fichier `05-lsgramcharac-model.xml` doit être amélioré. Pour l’instant, il n’est pas en mesure de prendre en charge la création de nouvelles caractéristiques grammaticales si celles-ci impliquent un niveau d’imbrication supérieur à 4 parents, qui est la limite actuelle (cas des classes phraséologiques et des caractéristiques d’autonomie phraséologique).

Évolution du modèle des caractéristiques grammaticales. À partir de la version 3.0 de l’export des données du RL-fr, on constate une réorganisation importante du modèle des caractéristiques grammaticales autour de la sous-hiérarchie **Statut lexical**, subordonnée à la classe **CARACTÉRISTIQUES GRAMMATICALES FONDAMENTALES**. Elle comprend elle-même quatre sous-classes :

1. **Statut phraséologique** → caractérisation en tant que *phrasème* ou en tant que constituant de phrasèmes, avec deux sous-classes :
 - (a) **Classe phraséologique** → pour les phrasèmes, caractérisation du degré de phraséologisation, selon la typologie de la Lexicologie Explicative et Combinatoire – par ex., **locution forte**, **semi-locution**, **locution faible**, **collocation**, ...
 - (b) **Autonomie phraséologique** → Pour les entités lexicales essentiellement collocatives : indique leur positionnement syntaxique relativement à leur base – par ex., **collocatif de son gouverneur syntaxique**.
2. **Statut sémantique** → dans la plupart des cas, il s’agit d’une classification des lexies qui sont sémantiquement des noms propres (en fonction de leur type de dénotation) – par ex., **anthroponyme**, **toponyme**, ... Voir le champ de commentaires pour chaque statut sémantique particulier.

Important ! Comme on le voit, le fait pour un nom d’être un nom propre est considéré comme relevant d’un statut sémantique et non d’une partie du

discours particulière. De ce fait, il n'existe plus qu'une unique partie du discours pour les lexèmes nominaux appelée **nom** (sous **Parties du discours > N**) : un nom sans statut sémantique associé appartient à la classe traditionnelle des « noms communs » ; un nom possédant un statut sémantique spécial appartient à la classe traditionnelle des « noms propres ».

3. **Statut pragmatique** → caractérisation du type de « pragmatité » associée à l'entité lexicale – par ex., **pragmatème** (Mel'čuk, 2020). Cette sous-classe de caractéristiques grammaticales demande à être théorisée.
4. **Statut langagier** → à mettre en rapport avec les caractéristiques sous **MARQUES D'USAGE > Marques d'usage langagier**, cette sous-classe concerne les constituants lexicaux de locutions qui ne sont pas de véritables sens lexicalisés dans la langue (contemporaine) – par ex., **lexie inactive**. Voir le champ de commentaires pour chaque statut langagier particulier.

Fichier 06-lsgramcharac-rel.csv

node	usagenote	usagenotevars	POS	phraseostruc	embeddedlex	othercharac	othercharacvars
ls:fr:node:30758			ls:fr:gc:21	NC Prép NC	((ls:fr:node:46192),(de), (ls:fr:node:46193,boucliers))	(ls:fr:gc:26,ls:fr:gc:166)	((,))
ls:fr:node:30811	(ls:fr:gc:6)	((,))	ls:fr:gc:21	NC Prép Art NC	((ls:fr:node:47539),(de),(le), (ls:fr:node:33810,))	(ls:fr:gc:28,ls:fr:gc:166, ls:fr:gc:98)	((,),(du culte))
ls:fr:node:30812	(ls:fr:gc:187,	((,))	ls:fr:gc:20			(ls:fr:gc:28)	((,))
ls:fr:node:30813			ls:fr:gc:24				
ls:fr:node:30814			ls:fr:gc:20			(ls:fr:gc:28)	((,))
ls:fr:node:30815	(ls:fr:gc:11)	((,))	ls:fr:gc:20			(ls:fr:gc:28,ls:fr:gc:47)	((,))

FIG. 6 : Extrait du fichier 06-lsgramcharac-rel.csv

node	token	ID du nœud lexical considéré.
usagenote	array of tokens	Liste des ID des marques d'usage (spéc , fam , etc.) du nœud lexical.
usagenotevars	array of arrays of tokens	Liste des variables éventuelles dans les noms des marques d'usage.
POS	token	ID de la partie du discours (<i>part of speech</i>).
phraseostruc	string	Chaîne de caractères identifiant le patron de structure lexico-syntaxique, pour une locution.
embeddedlex	array of arrays of tokens	Liste de couples (i) ID des lexies contenues dans la structure lexico-syntaxique, pour une locution et (ii) forme de ces lexies.
othercharac	array of tokens	ID des autres caractéristiques grammaticales.
othercharacvars	array of arrays of tokens	Liste des variables éventuelles dans les noms des caractéristiques grammaticales autres que les marques d'usage et les parties du discours.

TAB. 6 : Description des données du fichier 06-lsgramcharac-rel.csv

Comme on le voit, les structures lexico-syntaxiques des *locutions* (Pausé, 2017) sont intégrées dans le fichier 06-lsgramcharac-rel.csv. Ainsi, la première ligne de données de la Figure 6,

plus haut, indique que 「LEVÉE DE BOUCLIER」 (ID `ls:fr:node:30758`) est une locution nominale (ID `ls:fr:gc:21`) dont la structure lexico-syntaxique NC Prép NC comprend les deux lexies LEVÉE **IV** (ID `ls:fr:node:46192`) et BOUCLIER (ID `ls:fr:node:46193`).

Pour l’instant, l’export ne spécifie pas à quel nœud lexical de la structure lexico-syntaxique de la locution correspond chacune des lexies enchâssées de la colonne `embeddedlex`. Par exemple, dans l’exemple analysé ci-dessus, il n’est pas indiqué explicitement que LEVÉE **IV** est le premier nom commun⁶ (NC) de la chaîne NC Prép NC et BOUCLIER le second. De plus, la structure n’est pas encodée comme un arbre syntaxique véritable⁷, tel que représenté dans la Figure 7 :

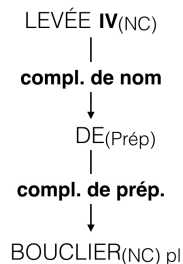


FIG. 7 : Structure lexico-syntaxique de la locution 「LEVÉE DE BOUCLIER」

Les listes de variables dans `usagenotevars` et `othercharacvars` peuvent être illustrées au moyen de la deuxième ligne de données de la Figure 6. Elle indique que 「MINISTRE DU CULTE」 (ID `ls:fr:node:30811`) possède trois caractéristiques grammaticales autres qu’une partie du discours et des marques d’usage. Le nom de la troisième de ces caractéristiques – `<<%>>` peut être omis – comporte une variable (%), qui est instanciée ici par *du culte*.

Vérifications effectuées à chaque mise à jour

- ⊙ Y a-t-il des lexies sans partie du discours ?
- ⊙ Y a-t-il des lexies associées à plusieurs parties du discours ?
- ⊙ Y a-t-il des lexies nominales sans genre grammatical ?
- ⊙ Y a-t-il des entités lexicales de nature syntagmatiques (donc, non lexémiques) sans statut phraséologique ?
- Y a-t-il des caractéristiques grammaticales du modèle non utilisées ?

7 Signifiants lexicaux

La morphologie des lexies est implantée de façon assez sophistiquée dans les Systèmes Lexicaux du type RL-fr au moyen de modèles morphologiques, qui permettent notamment de générer toutes les variantes flexionnelles d’une lexie. La stratégie lexicographique adoptée pour modéliser la morphologie des lexies est décrite dans Gader et al. (2014a).

Les règles morphologiques du modèle ne sont pas présentes dans l’export, où seuls figurent les *mots-formes* associés à chaque lexie et les traits morphologiques utilisés pour caractériser les mots-formes. Cette information est répartie en deux fichiers :

⁶Contrairement aux caractéristiques grammaticales (voir remarque *Important!*, p. 9), les patrons de structures lexico-syntaxiques font référence à deux classes distinctes de noms : nom commun (NC) et nom propre (NP).

⁷Ce type d’information sera disponible dans une version ultérieure de l’export.

- 07-lswordform-model.xml contient l'ensemble des traits morphologiques (*feature*), regroupés en catégories (*category*); il peut s'agir de grammèmes véritables (Singulier, Pluriel, ...) aussi bien que de paramètres d'usage (graphie rectifiée, forme élidée, ...).
- 08-lswordforms.csv contient tous les mots-formes de chaque lexie, chaque mot-forme étant associé à un ensemble de traits du fichier 07-lswordform-model.xml.

Fichier 07-lswordform-model.xml

```

<model>
<category id="ls:fr:flexc:1" name="Mode" pivot-name="mood">
<feature id="ls:fr:flexf:1" name="Indicatif" pivot-name="indicative" position="1"/>
<feature id="ls:fr:flexf:2" name="Conditionnel" pivot-name="conditional" position="2"/>
<feature id="ls:fr:flexf:3" name="Subjonctif" pivot-name="subjunctive" position="3"/>
<feature id="ls:fr:flexf:4" name="Impératif" pivot-name="imperative" position="4"/>
<feature id="ls:fr:flexf:5" name="Infinitif" pivot-name="infinitive" position="5"/>
<feature id="ls:fr:flexf:6" name="Participe" pivot-name="participle" position="6"/>
</category>
...
</model>

```

Figure 8: Extrait du fichier 07-lswordform-model.xml

id	token	ID d'un trait (grammatical, formel...) ou d'une catégorie de traits.
name	string	Nom d'un trait ou d'une catégorie de traits formulé dans la langue modélisée par le Système Lexical.
pivot-name	string	Nom d'un trait ou d'une catégorie de traits formulé au moyen d'un terme anglais pouvant servir de pivot dans le contexte de la mise en relation de plusieurs Systèmes Lexicaux (modélisant des langues différentes).
position	integer	Position absolue d'un trait dans une énumération de traits (<i>cf.</i> paramètre <i>features</i> du fichier 08-lswordforms.csv ci-dessous).

TAB. 7 : Description des données du fichier 07-lswordform-model.xml

Fichier 08-lswordforms.csv

node	features	signifier
ls:fr:node:26164	(ls:fr:flexf:1,ls:fr:flexf:10,ls:fr:flexf:15,ls:fr:flexf:16)	abandonna
ls:fr:node:26164	(ls:fr:flexf:1,ls:fr:flexf:10,ls:fr:flexf:13,ls:fr:flexf:16)	abandonnai
ls:fr:node:26164	(ls:fr:flexf:1,ls:fr:flexf:8,ls:fr:flexf:15,ls:fr:flexf:17)	abandonnaient

FIG. 9 : Extrait du fichier 08-lswordforms.csv

node	token	ID du mot-forme.
features	array of tokens	Liste ordonnée de façon normalisée des ID des traits (grammaticaux, formels, ...) portés par le mot-forme.
signifier	string	Signifiant du mot-forme.

TAB. 8 : Description des données du fichier 08-lswordforms.csv

Vérifications effectuées à chaque mise à jour

- ⊙ Y a-t-il des mots-formes associés à des règles morphologiques absentes du modèle?

8 Caractérisation sémantique

8.1 Étiquettes sémantiques

Les Systèmes Lexicaux utilisent une méthode de classification sémantique des lexies fondée sur une hiérarchie d'*étiquettes sémantiques*. Plus précisément, chaque étiquette est une instance d'une classe d'étiquettes, les classes étant organisées au sein d'une hiérarchie à héritage multiple. Ce système de classification sémantique est décrit dans Polguère (2011)⁸. Il réside dans deux fichiers de l'export :

- `09-lssemlabel-model.xml` contient la hiérarchie de classes d'étiquettes sémantiques avec leurs instances (c'est-à-dire, les étiquettes elles-mêmes) ;
- `10-lssemlabel-rel.csv` contient l'étiquetage des lexies individuelles.

⁸Voir aussi Milićević et Polguère (2010) pour ce qui est de l'héritage multiple comme encodage de l'*ambivalence sémantique* de certaines lexies.

Fichier 09-lssemlabel-model.xml

```
<model>
  <class id="ls:fr:semLabcl:534" name="QQCH." status="1" semfield="0" inheritancetype="0" comment="">
    <class id="ls:fr:semLabcl:253" name="ENTITÉ" status="1" semfield="0" inheritancetype="0" comment="">
      ...
    <instance id="ls:fr:semLab:281" name="entité" status="1" derivation="---" acttype="1" comment=""/>
  </class>
  <class id="ls:fr:semLabcl:273" name="FAIT" status="1" semfield="0" inheritancetype="0" comment="">
    <class id="ls:fr:semLabcl:35" name="ACTION" status="1" semfield="0" inheritancetype="0"
      comment="Fait ponctuel dont le premier actant est agent -> X est causateur volontaire">
      <class id="ls:fr:semLabcl:27" name="ACTE_DE_COMMUNICATION" status="1" semfield="1"
        inheritancetype="0" comment="">
        <class id="ls:fr:semLabcl:159" name="COMMUNICATION_LANGAGIÈRE" status="1"
          semfield="1" inheritancetype="0" comment="">
          ...
          <instance id="ls:fr:semLab:177" name="communication langagière" status="1"
            derivation="---" acttype="0" comment=""/>
          <instance id="ls:fr:semLab:178" name="dire [qqch.]" status="1"
            derivation="V0" acttype="0" comment=""/>
          <instance id="ls:fr:semLab:179"
            name="qui exprime un énoncé d'un certain type" status="1"
            derivation="A1" acttype="0" comment=""/>
          <instance id="ls:fr:semLab:974"
            name="relatif à une communication langagière" status="1" derivation="A0"
            acttype="0" comment=""/>
        </class>
        ...
        <instance id="ls:fr:semLab:27" name="acte de communication" status="1"
          derivation="---" acttype="0" comment=""/>
        <instance id="ls:fr:semLab:885" name="communiquer [qqch.]" status="1"
          derivation="V0" acttype="0" comment=""/>
      </class>
      ...
      <instance id="ls:fr:semLab:37" name="action" status="1" derivation="---" acttype="1" comment=""/>
      <instance id="ls:fr:semLab:926" name="faire qqch." status="1" derivation="V0" acttype="0"
        comment=""/>
    </class>
    ...
    <instance id="ls:fr:semLab:303" name="fait" status="1" derivation="---" acttype="1" comment=""/>
    <instance id="ls:fr:semLab:793" name="avoir lieu" status="1" derivation="V0" acttype="0"
      comment="Étiquette notamment de verbes supports de type Func0."/>
    <instance id="ls:fr:semLab:1028" name="commencer à avoir lieu" status="1"
      derivation="IncepV0" acttype="0" comment=""/>
  </class>
  <instance id="ls:fr:semLab:595" name="qqch." status="1" derivation="---" acttype="1" comment=""/>
</class>
</model>
```

Figure 10: Extrait du fichier 09-lssemlabel-model.xml

class	id	token	ID de la classe d'étiquettes sémantiques.
	name	string	Nom de la classe.
	status	integer	Statut lexicographique de la classe. Deux valeurs sont utilisées : 0 = classe valide <i>vs</i> 1 = classe non encore validée.
	semfield	boolean	Indique si la classe est utilisable pour identifier un champ sémantique du modèle.
	inheritancetype	integer	Type de connexion à la classe-mère considérée. Trois valeurs sont utilisées : 0 = héritage simple (une seule mère), 1 = héritage multiple inclusif (<i>and/or</i>), 2 = héritage multiple exclusif (<i>or</i>).
	comment	string	Commentaires éventuels sur la classe.
instance	id	token	ID de l'instance de classe = de l'étiquette.
	name	string	Nom de l'étiquette.
	status	integer	Statut lexicographique de l'étiquette – mêmes valeurs que pour les statuts de classes plus haut.
	derivation	string	Nature sémantico-grammaticale de l'étiquette, exprimée en termes de <i>fonctions lexicales</i> – voir explications plus bas.
	acttype	boolean	Indique si l'étiquette peut être utilisée pour typer une variable actancielle dans une définition lexicographique.
	comment	string	Commentaires éventuels sur l'étiquette.

TAB. 9 : Description des données du fichier `09-lssemlabel-model.xml`

Tel qu'expliqué en détail dans Polguère (2011), une étiquette sémantique type est nominale et est utilisable uniquement pour étiqueter des lexies nominales. Chaque classe d'étiquette, cependant, regroupe potentiellement une grappe d'étiquettes sémantiques qui permettent d'étiqueter des dérivés sémantiques (verbaux, adjectivaux, ... et même nominaux) des lexies étiquetées par l'étiquette nominale « de base » de la classe. Le paramètre **derivation** associé à chaque étiquette spécifie son statut (de base ou dérivational) dans la classe. Le Tableau 10 donne les valeurs possibles de ce paramètre, dans l'état courant du modèle lexicologique.

---	Étiquette de base de la classe d'étiquettes pour des lexies nominales.
Convij	Conversif de ---.
V0	Verbalisation de ---.
IncepV0	Inchoatif de la verbalisation de ---.
FinV0	Terminatif de la verbalisation de ---.
V0Convij	Verbalisation conversive de ---.
A0	Adjectivation de ---.
Adv0	Adverbalisation de ---.
S1	Nom de 1 ^{er} actant de ---.
S2	Nom de 2 ^e actant de ---.
S3	Nom de 3 ^e actant de ---.
A1	Adjectif qui dénote le 1 ^{er} actant de ---.
A2	Adjectif qui dénote le 2 ^e actant de ---.
A3	Adjectif qui dénote le 3 ^e actant de ---.
Adv1	Adverbe qui signifie '[faire qqch.] en ---'.
A2Manif	Adjectif qui dénote N dans --- <i>se manifeste dans N</i> .
Able1	Adjectif qui dénote un 1 ^{er} actant possible de ---.
Able2	Adjectif qui dénote un 2 ^e actant possible de ---.
Caus	Verbe qui signifie 'causer ---'.
ElementOf	Nom qui dénote un élément de ---.
SetOf	Nom qui dénote un ensemble de ---.
PartOf	Nom qui dénote une partie de ---.
Small	Nom qui dénote un petit ---.
Type	Nom qui dénote un type de ---.
Claus	Énoncé (= clausatif) « --- ».
Incep	Inchoatif de ---.
Fin	Terminatif de ---.

TAB. 10 : Valeurs possibles du paramètre `derivation`

La structure XML du fichier `09-lssemlabel-model.xml` encode une hiérarchie de classes d'étiquettes sémantiques dont un échantillon est visualisé dans la Figure 11⁹.

⁹Les instances de chaque classe d'étiquettes apparaissent dans une boîte à côté du nom de la classe.

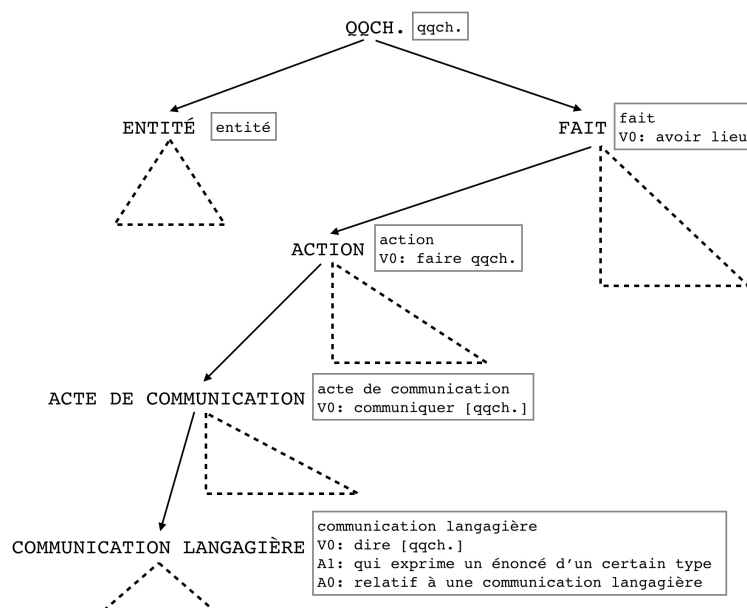


FIG. 11 : Échantillon de la hiérarchie des classes d'étiquettes sémantiques

Fichier 10-lssemlabel-rel.csv

node	label	%
ls:fr:node:26162	ls:fr:semlab:152	60
ls:fr:node:26164	ls:fr:semlab:734	60
ls:fr:node:26166	ls:fr:semlab:699	100
ls:fr:node:26167	ls:fr:semlab:583	60

FIG. 12 : Extrait du fichier 10-lssemlabel-rel.csv

node	token	ID de la lexie étiquetée.
label	token	ID de l'étiquette de la lexie.
%	integer	Degré de confiance de l'information.

TAB. 11 : Description des données du fichier 10-lssemlabel-rel.csv

Vérifications effectuées à chaque mise à jour

- Y a-t-il des étiquettes non utilisées ?

8.2 Formes propositionnelles

Les *formes propositionnelles* associées à chaque lexie sont entrées avec l'éditeur lexicographique sous le contrôle d'un modèle hiérarchique des patrons de formes propositionnelles admissibles. Cela permet d'assurer une certaine systématisme dans la description des *structures actancielles* lexicales dans l'ensemble du Système Lexical.

L'export ne contient pas le modèle de patrons de formes propositionnelles, mais un unique fichier – 11-lspropform-rel.csv – qui rassemble l'ensemble de toutes les associations « lexie-forme propositionnelle ».

Fichier 11-lspropform-rel.csv

node	propform	tildevalue	%	actantslist
ls:fr:node:26162	[\$1] ~ \$2		100	(\$1=X,\$2=Y)
ls:fr:node:26163	[\$1] ~ de \$2		100	(\$1=X,\$2=Y)
ls:fr:node:26166	~		100	∅
ls:fr:node:26167	\$1 ~ \$2	abîme	100	(\$1=X,\$2=Y)
ls:fr:node:26168	\$1 ~ \$2 à \$3	abonne	100	(\$1=X,\$2=Y,\$3=Z)

FIG. 13 : Extrait du fichier 11-lspropform-rel.csv

node	token	ID du nœud lexical considéré (la lexie vedette d'un article lexicographique).
propform	string	Forme propositionnelle proprement dite, où le tilde (~) indique la position de la lexie vedette dans la formule et \$1, \$2, ... indiquent les positions actancielles pour le 1 ^{er} , 2 ^e , ... actant.
tildevalue	string	Optionnel : forme devant être utilisée à la place du nom lexicographique de la lexie vedette pour instancier le tilde de la formule dans la présentation « textuelle » de l'information lexicale, au sein d'un article lexicographique.
%	integer	Degré de confiance de l'information.
actantslist	array of tokens	Liste des variables X, Y, ... devant être utilisées à la place de \$1, \$2, ... dans la présentation de l'information lexicale au sein d'un article lexicographique.

TAB. 12 : Description des données du fichier 11-lspropform-rel.csv

Ainsi, selon les explications données ci-dessus, la version article lexicographique de la dernière ligne de la Figure 13 se formulera de la façon suivante : X abonne Y à Z.

Vérifications effectuées à chaque mise à jour

- Y a-t-il des espaces non valides dans les formes propositionnelles ?
- ⊙ Y a-t-il des formes propositionnelles dans lesquelles le même nom de variable (X, Y, ...) est utilisé pour deux positions actancielles différentes ?

8.3 Définitions

L'écriture de définitions lexicales (Mel'čuk et Polguère, 2016, 2018) en est à un stade embryonnaire dans le RL-fr et seules quelques dizaines de définitions sont encodées dans l'export courant¹⁰. Cet encodage se fait à travers deux fichiers :

- 12-definition.dtd contient la description de la structure des *definiens* (= paraphrases définitionnelles) des définitions lexicales ;
- 13-lsdef.csv contient les *definiens* des définitions lexicales encodées en format XML et HTML.

Nous ne commentons ici que le second fichier.

¹⁰La plupart de ces définitions (une centaine) relèvent de la terminologie fondamentale de la chimie (Mikhel, 2022).

Fichier 13-lsdef.csv

node	token	ID du nœud lexical considéré (la lexie vedette d'un article lexicographique).
def_XML	string	Paraphrase définitionnelle dans un format XML.
def_HTML	string	Paraphrase définitionnelle dans un format HTML.

TAB. 13 : Description des données du fichier 13-lsdef.csv

Tel qu'expliqué dans Mel'čuk et Polguère (2016, sec. 2.2.3), les paraphrases définitionnelles, qui sont présentées sous forme textuelles dans les dictionnaires, correspondent formellement à des *réseaux sémantiques*. Le champ `def_XML` contient à la fois la paraphrase textuelle et un balisage XML qui encode la structure de réseau sémantique sous-jacente – cf. Mel'čuk et Polguère (2016, pp. 77–78).

Le champ `def_HTML` donne la version textuelle de la définition encodée en HTML (avec hyperliens) pour des visualisations textuelles, comme celles du navigateur Spiderlex¹¹.

Vérifications effectuées à chaque mise à jour

- ⊙ Y a-t-il des sémantèmes utilisés dans les définitions qui ne sont pas répertoriés dans la table de sémantèmes ?

9 Liens de fonctions lexicales (FL)

Pour une introduction succincte à la notion de *fonction lexicale* – désormais *FL* –, on pourra se reporter à Polguère (2016), pp. 195–210 ; la présentation détaillée à jour des FL est faite dans Mel'čuk et Polguère (2021).

Note. La version du modèle des FL prise en compte dans le RL-fr est celle de Mel'čuk et Polguère (2021).

Les liens de FL tissent l'ossature des Systèmes Lexicaux. Ils résident dans deux fichiers de l'export :

- `14-lslf-model.xml` contient le modèle hiérarchique des FL : chaque FL appartient à une « famille » et chaque famille est elle-même élément d'un « groupe » de familles ;
- `15-lslf-rel.csv` contient l'ensemble de liens de FL entre lexies individuelles.

¹¹<https://lexical-systems.atilf.fr/consultation-spiderlex/>

Fichier 14-lslf-model.xml

```
<model>
  ...
  <group>
    <family id="ls:fr:lffam:3" name="Syn">
      <lexicalfunction id="ls:fr:lf:2" name="Syn" status="0"
        linktype="paradigmatic" standardness="simple standard" semantics="2">
        <![CDATA[
          <span class ="FLb">Syn</span>
        ]]>
      </lexicalfunction>
      ...
    </family>
    ...
  </group>
  ...
  <group>
    <family id="ls:fr:lffam:15" name="S1">
      <lexicalfunction id="ls:fr:lf:31" name="S_1" status="0"
        linktype="paradigmatic" standardness="simple standard" semantics="2">
        <![CDATA[
          <span class ="FLb">S</span> <span class ="sub">1</span>
        ]]>
      </lexicalfunction>
      ...
    </family>
    ...
  </group>
  ...
</model>
```

Figure 14: Extrait du fichier 14-lslf-model.xml

id	token	ID de la fonction lexicale ou de la famille de fonctions lexicales.
name	string	Nom de la fonction lexicale ou de la famille de fonctions lexicales. Les caractères « _ » et « ^ » introduisent, respectivement, un indice <i>vs</i> un exposant – par exemple, S_1 pour la fonction lexicale S₁ .
status	integer	Statut lexicographique de la fonction lexicale. Deux valeurs sont utilisées : 0 = fonction lexicale valide <i>vs</i> 1 = fonction lexicale non encore validée.
linktype	token	Type de la fonction lexicale relativement à la nature sémantique ou combinatoire du lien qu'elle tisse. Deux valeurs sont utilisées : paradigmatic <i>vs</i> syntagmatic .
standardness	string	Type de la fonction lexicale relativement à son caractère formel : simple ou complexe, plus ou moins standard, etc.
semantics	integer	Poids sémantique des liens lexicaux tissés par la fonction lexicale : 0 = poids nul, 1 = poids intermédiaire et 2 = poids maximal.
CDATA	string	Variante formatée du nom de fonction lexicale (paramètre name ci-dessus), pour affichage dans un article lexicographique.

TAB. 14 : Description des données du fichier `14-lslf-model.xml`

À noter à propos de l'exportation des données dans le fichier `14-lslf-model.xml` :

- les éventuelles familles de FL qui ne regroupent aucune FL dans le RL-fr ne sont pas intégrées à l'export ;
- dans les noms de FL, les occurrences du caractère « & » ont été remplacées par des « & ».

Rappelons qu'une fonction lexicale donnée **f** s'applique aux lexies L de la langue – **f(L)** – pour retourner une valeur {V} qui est un ensemble de lexies (ou expressions linguistiques) exprimant le contenu sémantico-syntaxique de **f** relativement à L. Cela se représente par la formule suivante, qui est une métaphore formelle de l'application d'une fonction mathématique :

$$\mathbf{f}(L) = \{V\}.$$

Pour une illustration, voir la formule (1) plus bas, p. 22.

Fichier 15-lslf-rel.csv

source	lf	target	form	separator	merged	syntacticframe	constraint	position
ls:fr:node:35193	ls:fr:lf:5	ls:fr:node:28895		,	0			1
ls:fr:node:35193	ls:fr:lf:5	ls:fr:node:35202		,	0			2
ls:fr:node:35193	ls:fr:lf:524	ls:fr:node:26369		,	0			1
ls:fr:node:35193	ls:fr:lf:524	ls:fr:node:26520		,	0			2
ls:fr:node:35193	ls:fr:lf:524	ls:fr:node:54368		;	0			3
ls:fr:node:35193	ls:fr:lf:524	ls:fr:node:26980		;	0			4
ls:fr:node:35193	ls:fr:lf:524	ls:fr:node:28171		;	0			5
ls:fr:node:35193	ls:fr:lf:74	ls:fr:node:30608		,	0	de ~		1
ls:fr:node:35193	ls:fr:lf:74	ls:fr:node:33239		,	0	de ~		2
ls:fr:node:35193	ls:fr:lf:74	ls:fr:node:30335		;	0	de ~		3
ls:fr:node:35193	ls:fr:lf:228	ls:fr:node:28235		,	0	ART ~		1
ls:fr:node:35193	ls:fr:lf:228	ls:fr:node:47505		,	0	ART ~		2

FIG. 15 : Extrait du fichier 15-lslf-rel.csv

source	token	ID de la lexie source du lien de FL.
lf	token	ID de la FL.
target	token	ID de la cible (généralement une lexie) du lien de FL.
form	string	Optionnel : formulation à substituer au nom lexicographique de la cible du lien dans un article lexicographique.
separator	token	Séparateur entre cette cible et l'éventuelle cible précédente dans la liste des éléments de la valeur de l'application $\mathbf{f}(L)$. Les séparateurs possibles sont : « , » (faible distance sémantique entre cette cible et la précédente), « ; » (distance sémantique assez notable) et « < » (distance sémantique correspondant à une augmentation d'amplitude, intensité, ...).
merged	boolean	Booléen qui indique s'il s'agit d'une valeur non fusionnée (0) – cas par défaut – ou fusionnée (1), dans le cas d'une FL syntagmatique.
syntacticframe	string	Optionnel : régime syntaxique de la cible du lien de FL.
constraint	string	Optionnel : contraintes grammaticales, stylistiques, etc.
position	integer	Position relative de la cible dans la liste des cibles pour la même application $\mathbf{f}(L)$.

TAB. 15 : Description des données du fichier 15-lslf-rel.csv

Les liens de FL connectent des lexies individuelles, mais dans les faits, toute application de FL \mathbf{f} à une lexie donnée L – c'est-à-dire $\mathbf{f}(L)$ – retourne une valeur qui est un **ensemble** de lexies. Il est donc nécessaire, pour chaque lien, d'indiquer comment la cible se positionne à l'intérieur de la liste des éléments de la valeur de $\mathbf{f}(L)$ – cf. valeur de **position**.

Par exemple, les lignes de données 8 à 10 de la Figure 15, rendent compte de l'application de la FL **Sing** (ID `ls:fr:lf:74`) à la lexie SOUPE (ID `ls:fr:node:35193`). Elles se visualiseront dans un article lexicographique de la façon suivante :

(1) **Sing**(*soupe*) = *assiette*¹ **I.b** [*de ~*], *assiettée* [*de ~*] ; *bol*¹ **I.b** [*de ~*].

Vérifications effectuées à chaque mise à jour

- Y a-t-il des FL du modèle non utilisées ?

10 Exemples lexicographiques

Les exemples lexicographiques sont introduits comme des entités informationnelles autonomes dans les Systèmes Lexicaux. L'exemplification des lexies se fait par association des lexies individuelles à un ou plusieurs exemples – voir Lux-Pogodalla (2014).

L'information sur l'exemplification réside dans trois fichiers de l'export :

- 16-lsexsource-model.xml contient le modèle des sources des exemples lexicographiques (notamment, les corpus textuels) ;
- 17-lsex.csv contient les exemples lexicographiques proprement dits ;
- 18-lsex-rel.csv contient les associations lexie → exemple.

Fichier 16-lsexsource-model.xml

```
<model>
  ...
  <section id="ls:fr:exsect:3" name="Citations de corpus journalistiques hors Frantext">
    <source id="ls:fr:exsrc:3" name="L'Est Républicain"/>
    <source id="ls:fr:exsrc:4" name="Chambers-Rostand"/>
  </section>
  ...
</model>
```

Figure 16: Extrait du fichier 16-lsexsource-model.xml

id	token	ID de corpus ou de section dans l'ensemble des corpus (= type de corpus).
name	string	Nom du corpus ou de la section.

TAB. 16 : Description des données du fichier 16-lsexsource-model.xml

Fichier 17-lsex.csv

id	source	status	content	title	authors	location	date
ls:fr:ex:13	ls:fr:exsrc:1	1	<html><body><p>Nous voulions kidnapper le gérant et le juger. Nous avons entamé une campagne de propagande contre « le gérant fasciste de Seyssinet ».</p></body></html>	Voyage au bout de la révolution : de Pékin à Sochaux	Brière-Blanchet, Claire	p. 116	2009
ls:fr:ex:16	ls:fr:exsrc:2	1	<html><body><p>Les ravisseurs ont profité d'une absence de vigilance de la mère pour kidnapper la petite Amanda.</p></body></html>			http://www.yozone.fr/spip.php?article4460	02/2008
ls:fr:ex:135	ls:fr:exsrc:5	1	<html><body><p>Quoi de mieux que de finir la semaine par un petit basket entre midi et 2 ?! Alors venez mettre quelques paniers, avec de la technique, du style, ou simplement de la chance !</p></body></html>			http://www.savoie-technolac.com/acs/10/127-acsculture-sport.htm	16/10/2012

FIG. 17 : Extrait du fichier 17-lsex.csv

id	token	ID de l'exemple.
source	token	ID de la source.
status	integer	Statut lexicographique de l'exemple. Deux valeurs sont utilisées : 0 = exemple valide <i>vs</i> 1 = exemple non encore validé.
content	string	Contenu textuel de l'exemple. Celui-ci est stocké en HTML, car il peut contenir du formatage (italiques, gras, mises en exposant, etc.).
title	string	Titre de l'ouvrage (ou assimilé) d'où provient l'exemple.
authors	string	Auteur(s) de l'exemple.
location	string	Localisation de la publication (éditeur, page web, etc.).
date	token	Date associée à chaque exemple. La spécification de la date est scindée en trois champs dans la base lexicographique du Système Lexical ; elle occupe ici un seul champ, sous la forme AAAA, MM/AAAA ou JJ/MM/AAAA.

TAB. 17 : Description des données du fichier 17-lsex.csv

Note. Seuls les exemples associés à au moins un nœud du graphe ont été exportés.

Fichier 18-lsex-rel.csv

node	example	occurrence	position	%
ls:fr:node:44432	ls:fr:ex:13882	163,172;	0	100
ls:fr:node:44432	ls:fr:ex:13746	66,76;	1	100
ls:fr:node:44432	ls:fr:ex:13779	87,96;	2	100
ls:fr:node:44432	ls:fr:ex:13804	16,25;	3	100
ls:fr:node:44432	ls:fr:ex:13859	302,311;	4	100
ls:fr:node:44433	ls:fr:ex:13750	120,129;	0	100
ls:fr:node:44433	ls:fr:ex:13821	197,207;273,283;	1	100

FIG. 18 : Extrait du fichier 18-lsex-rel.csv

node	token	ID de la lexie exemplifiée.
example	token	ID d'un exemple illustratif associé à la lexie.
occurrence	string	Segment(s) de texte de l'exemple où apparaît la lexie. Chaque segment est indiqué par la position dans l'exemple du premier caractère du segment, suivie du caractère « , », lui-même suivi de la position du dernier caractère. Les segments sont séparés par le caractère « ; ».
position	integer	Position relative de l'exemple, parmi les autres exemples, dans une visualisation de type article lexicographique.
%	integer	Degré de confiance de l'information.

TAB. 18 : Description des données du fichier 18-lsex-rel.csv

Le paramètre **occurrence** contient plus d'un segment dans un des deux cas suivants :

- l'exemple contient plusieurs occurrences de la lexie qu'il sert à illustrer – par exemple, *Être ou ne pas être*, dans un exemple illustrant le lexème verbal ÊTRE_(V) 1 ;
- la lexie est exprimée sous la forme d'un syntagme disjoint – par exemple, *avait maladroitement renversé*, dans un exemple illustrant le lexème RENVERSER 1.1, ou *jouer souvent des coudes*, dans un exemple illustrant la locution 「JOUER DES COUDES」.

Vérifications effectuées à chaque mise à jour

Les exemples lexicographiques du RL-fr sont également distribués sous la forme d'un corpus indépendant appelé *BEL-RL-fr*. Il est téléchargeable sur la plateforme Ortolang à l'adresse suivante :

<https://hdl.handle.net/11403/examples-ls-fr>.

Cette base d'exemple est mise à jour à chaque mise à jour de la distribution des systèmes lexicaux. Se reporter à la document du BEL-RL-fr pour la liste des vérifications effectuées sur les exemples.

Références

- Nabil Gader, Aurore Koehl et Alain Polguère. A Lexical Network with a Morphological Model in It. Dans *Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon (CogALex IV)*, pages 154–165, Dublin, 2014a. Association for Computational Linguistics and Dublin City University.
- Nabil Gader, Veronika Lux-Pogodalla et Alain Polguère. Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor. Dans *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CogALex III)*, pages 109–125, Mumbai, 2012. The COLING 2012 Organizing Committee.
- Nabil Gader, Sandrine Ollinger et Alain Polguère. One Lexicon, Two Structures : So What Gives ? Dans Heili Orav, Christiane D. Fellbaum et Piek Vossen, dir., *Proceedings of the Seventh Global Wordnet Conference (GWC2014)*, pages 163–171, Tartu (Estonie), 2014b. Global WordNet Association.
- Veronika Lux-Pogodalla. Intégration relationnelle des exemples lexicographiques dans un réseau lexical. Dans Brigitte Bigi, dir., *Actes de TALN 2014*, pages 586–591, Marseille, 2014. Laboratoire Parole et Langage, Aix-en-Provence.
- Veronika Lux-Pogodalla et Alain Polguère. Construction of a French Lexical Network : Methodological Issues. Dans *Proceedings of the First International Workshop on Lexical Resources, WoLeR 2011. An ESSLLI 2011 Workshop*, pages 54–61, Ljubljana, Slovenia, 2011.
- Igor Mel'čuk. Phraseology in the Language, in the Dictionary, and in the Computer. Dans Koenraad Kuiper, dir., *Yearbook of Phraseology*, volume 3, pages 31–56. De Gruyter Mouton, Berlin, 2012.
- Igor Mel'čuk. Clichés and pragmatemes. *Neophilologica*, 32 :9–21, 2020.
- Igor Mel'čuk et Alain Polguère. La définition lexicographique selon la Lexicologie Explicative et Combinatoire. *Cahiers de lexicologie*, 109 :61–91, 2016.
- Igor Mel'čuk et Alain Polguère. Theory and Practice of Lexicographic Definition. *Journal of Cognitive Science*, 19(4) :417–470, 2018. URL <http://cogsci.snu.ac.kr/jcs/issue/vol19/no4/01+Igor+Alain.pdf>.
- Igor Mel'čuk et Alain Polguère. Les fonctions lexicales dernier cri. Dans Sébastien Marengo, dir., *La Théorie Sens-Texte. Concepts-clés et applications*, Dixit Grammatica, pages 75–155. L'Harmattan, Paris, 2021.
- Polina Mikhel. *Multilingual study of chemistry lexicon at the interface of terminology and general language*. Thèse de doctorat, Université de Lorraine, Nancy, 2022.
- Jasmina Milićević et Alain Polguère. Ambivalence sémantique des noms de communication langagière du français. Dans F. Neveu, V. Muni Toke, J. Durand, T. Klingler, L. Mondada et S. Prévost, dir., *Actes de la section « Lexique et morphologie » du 2^e Congrès Mondial de Linguistique Française (CMLF'10)*, pages 1029–1050, La Nouvelle-Orléans, 12–15 juillet 2010. Institut de Linguistique Française (ILF).

- Marie-Sophie Pausé. *Structure lexico-syntaxique des locutions du français et incidence sur leur combinatoire*. Thèse de doctorat, Université de Lorraine, Nancy, 2017.
- Alain Polguère. Classification sémantique des lexies fondée sur le paraphrasage. *Cahiers de lexicologie*, 98 :197–211, 2011.
- Alain Polguère. From Writing Dictionaries to Weaving Lexical Networks. *International Journal of Lexicography*, 27(4) :396–418, 2014.
- Alain Polguère. *Lexicologie et sémantique lexicale. Notions fondamentales*. Paramètres. Les Presses de l'Université de Montréal, Montréal, 3^e édition, 2016.
- Alain Polguère. A Lexicographic Approach to the Study of Copolysemy Relations. *Russian Journal of Linguistics*, 22(4) :788–820, 2018. URL <http://journals.rudn.ru/linguistics/article/view/20165>.