

Mémo RLF

Exemples lexicographiques

Jean-Luc Benoit[†], Veronika Lux-Pogodalla[†] et Alain Polguère[‡]

[†]ATILF CNRS, [‡]Nancy-Université & ATILF CNRS

27 novembre 2011

Table des matières

1	Principes de sélection des exemples lexicographiques	1
2	Structure de la zone d'exemples d'un article de RLF	3
2.1	Citations du corpus Frantext (obligatoires)	3
2.2	Citations du corpus FrWac (obligatoires)	3
2.3	Citations de corpus journalistiques hors Frantext (obligatoire)	4
2.4	Citations hors bases	5
2.5	Exemples fabriqués ou adaptés	6
2.6	Citations hors-norme	7
3	À propos des corpus oraux	7
	Remerciements	7
	Bibliographie	7

Résumé

Nous présentons la structure générale de la zone d'exemples d'un article de RLF. Il s'agit principalement d'identifier les différentes classes d'exemples potentiellement présents dans les articles et d'indiquer comment ces exemples doivent être sélectionnés et présentés par le lexicographe.

1 Principes de sélection des exemples lexicographiques

Un *exemple lexicographique*¹ est introduit dans la *zone d'exemples* d'un article de lexie – le *mot-vedette* – car :

- il contient au moins une occurrence d'emploi du mot-vedette ;
- il permet d'illustrer une ou plusieurs propriétés (sémantiques, formelles, combinatoires) du mot-vedette.

Sept principes fondamentaux contrôlent la sélection des exemples dans le RLF.

1. L'exemple doit permettre de circonscrire avec suffisamment de précision le sens du mot-vedette, tout particulièrement lorsque ce dernier appartient à un vocable polysémique.

1. Nous dirons dorénavant simplement *exemple*, sauf si le contexte requiert d'être plus explicite.

2. Autant que possible, un exemple est une phrase syntaxique (1a). Il peut cependant s'agir d'un énoncé non phrastique, comme un syntagme nominal (1b) ou prépositionnel (1c), ou d'une suite d'énoncés formant un très court texte (1d).

Important Les exemples apparaissent ici dans une police sans empattement (angl. *sans serif*), plutôt qu'en italique. Chaque exemple contient en gras le mot-clé dont il illustre l'emploi.

- (1) a. Franchomme, lui, **comptait** ses œufs, ses noix et ses pommes.
[PERRY Jacques, *Vie d'un païen*, 1965, p. 31]
b. La pétition bidon des **vrais-faux** riches
[*Le Canard enchaîné*, 31 août 2011, p. 1, titre]
c. D'une **guerre** deux coups
[*Le Canard enchaîné*, 31 août 2011, p. 8, titre]
d. Ni commode, ni table, ni chaise, ni bergère autour de nous, ne restaient plus que le lit, qui est un droit légal, et, comme un crapaud sur la moquette jaune, le téléphone. . .
– On vous **a cambriolé!** criai-je.
[BILLETDOUX Marie, *Un peu de désir sinon je meurs*, 2006, p. 108]

3. Si l'exemple est une citation, il doit provenir d'un texte qui n'est pas antérieur à 1950.
4. Les troncations de phrases ou de syntagmes – cf. (2) ci-dessous (extrait du TLFi) –, sont totalement proscrites.

- (2) Le bâtiment (...) évoquait l'idée d'un bateau de plaisance (...) assurément capable de tenir la haute mer (...). Le **tableau** d'arrière ne portait ni nom ni trace aucune de l'immatriculation nautique.
[GRACQ Julien, *Le rivage des Syrtes*, 1951, p. 77]

5. De façon plus générale, une citation ne doit jamais être altérée, même pour en corriger les erreurs de français. À l'exception des cas de normalisation de certains caractères (voir prochaine consigne), la citation n'est une citation que si elle est dénuée de toute intervention linguistique du lexicographe. Si l'on veut absolument introduire une citation dans laquelle apparaît une erreur de français, on fera suivre cette dernière de la mention « [sic] », par exemple :

- (3) C'est aux Billard (les neveux **nunuche** [sic] de Tatie Danielle) qu'elle s'identifie.
[*Le Monde*, 12 avril 1990]

Noter que le *sic* du RLF, comme celui du TLF, s'écrit entre crochets et non entre parenthèses. Cela permet de mieux mettre en évidence le fait que la remarque en question est le fait du lexicographe, une citation pouvant tout à fait contenir un (*sic*).

6. Il faut veiller à la normalisation systématique des caractères suivants :
– apostrophe courbe (', et non ');
– guillemets dits *français* (« », et non “ ”);
– espaces insécables introduites selon les normes typographiques du français de France. Il faut aussi systématiquement rétablir les accents sur les majuscules, lorsque ceux-ci ont été omis.
7. Les exemples doivent être classés, selon leur type, dans des sections spécifiques de la zone d'exemples, tel que détaillé dans la section suivante. Seuls les exemples classables dans les sections prévues peuvent être retenus.

2 Structure de la zone d'exemples d'un article de RLF

La zone d'exemples d'un article de RLF est structurée en six sections, correspondant aux six classes suivantes d'exemples :

1. citations du corpus Frantext ;
2. citations du corpus FrWac ;
3. citations de corpus journalistiques hors Frantext ;
4. citations de corpus oraux TCOF ;
5. citations hors bases ;
6. exemples fabriqués ou adaptés ;
7. citations hors-norme.

Seules les trois premières sections de la zone d'exemples doivent nécessairement être complétées, les autres sections correspondant à des types d'exemples dont la présence dans un article de RLF complet est optionnelle.

Attention ! Lorsque nous indiquons ci-dessous qu'une section d'exemples est obligatoire, il faut interpréter le terme *obligatoire* avec une certaine souplesse. Il signifie qu'il faut obligatoirement chercher à compléter la section en question ; si aucun exemple n'apparaît pour une classe « obligatoire », cela signifie qu'il est tout simplement impossible de trouver un tel exemple pour la classe en question (par exemple, pour une acception de vocable relevant plutôt de la langue parlée que l'on ne peut retrouver dans Frantext).

2.1 Citations du corpus Frantext (obligatoires)

Le corpus *Frantext* (Bernard et al., 2002) est interrogeable en ligne² avec l'interface Stella. Il doit être systématiquement consulté, notamment pour les textes littéraires qu'il contient. Il faut cependant veiller à respecter la consigne 3 (page 2, ci-dessus) et, donc, n'utiliser que des citations de textes dont la publication n'est pas antérieure à 1950³.

Le formatage d'une citation – italique, gras et saut de ligne – doit toujours être restitué.

Le référencement d'une citation de Frantext est mis entre crochets, à la ligne, selon le patron suivant :

[<NOM DE FAMILLE DE L'AUTEUR> <Prénom>, <Titre de l'ouvrage>, <Date de publication>, p. <numéro de page>]

Par exemple :

- (4) À midi, et le soir vers six heures, il s'élevait de là une sorte de **mugissement** plaintif : c'est la sirène, disait la mère, comme elle aurait dit l'heure.
[SIZUN Marie, *Éclats d'enfance*, 2009, p. 25]

Remarque En cas d'indication de plusieurs dates dans Frantext, retenir l'édition la plus ancienne.

2.2 Citations du corpus FrWac (obligatoires)

Le corpus *FrWac* – construit dans le contexte du projet *WaCky wide web* (Baroni et al., 2009)⁴ – est une image du contenu des pages Internet du domaine **.fr** construite vers 2009 et

2. <http://www.frantext.fr>

3. Nous avons actuellement accès à 905 textes, correspondant à plus de 62 millions de mots-occurrences.

4. Site du projet : <http://wacky.sslmit.unibo.it/doku.php>; en local, sur les serveurs de l'ATILF : <https://arcas.atilf.fr/cqweb/>.

comportant environ 1,6 milliard de mots. Il sera prochainement disponible dans une version enrichie, où chaque page sera associée à une note reflétant la qualité du français de la page en question (obtenue par un traitement automatique effectué au moyen du logiciel Antidote de Druide informatique⁵). Grâce à cette note, il sera possible de restreindre l'espace de recherche sur le corpus pour que le concordancier utilisé ne retourne que des contextes de bonne qualité linguistique, qui pourront servir de citations. Rappelons qu'une citation ne doit jamais être altérée, même pour en corriger les erreurs de français (consigne 5, page 2).

Le référencement d'une citation de FrWac est mis entre crochets, à la ligne, selon le patron suivant :

[FRWAC, <numéro de fichier>]

Par exemple :

- (5) Sur les terres afghanes, le pavot **fleurit** comme jamais : la production d'opium a encore doublé cette année, atteignant 8 400 tonnes, un record depuis la chute des talibans.
[FRWAC, 61691636]

2.3 Citations de corpus journalistiques hors Frantext (obligatoire)

Pour l'instant, nous n'avons qu'un corpus journalistique, celui de l'*Est Républicain* – 1999, 2002 et 2003⁶ – accessible par le concordancier *CQP-Web* (Hardy, à paraître)⁷. Nous devrions disposer prochainement d'autres corpus journalistiques, par l'entremise de MVS, notre partenaire industriel dans le projet RELIEF.

Le référencement d'une citation de corpus journalistique est mis entre crochets, à la ligne, selon le patron suivant :

[<Nom du journal>, <Date (jour, mois et année)>]

Par exemple :

- (6) Gabriel, 21 ans, venait de **cambrioler** la cave d'un retraité et d'y dérober 17 bouteilles.
[*L'Est Républicain*, 23 mai 2002]

Remarque Contrairement à la pratique adoptée dans le TLF, il ne faut pas utiliser les abréviations *janv.*, *fév.*, *avr.*, *sept.*, *oct.*, *nov.* et *déc.* Les noms de mois s'écrivent toujours au complet.

Si la citation est un titre, on indique :

[<Nom du journal>, <Date (jour, mois et année)>, titre]

Par exemple (reprise de (1b) plus haut) :

- (7) La pétition bidon des **vrais-faux** riches
[*Le Canard enchaîné*, 31 août 2011, titre]

Remarque Nous avons renoncé à mentionner les numéros de pages et de colonnes pour les citations journalistiques, car cette information n'est pas nécessairement disponible dans les versions électroniques des journaux.

5. <http://www.druide.com/antidote.html>

6. Le corpus de l'*Est Républicain* contient toutes les éditions régionales et locales pour 1999, 2002, et 2003. Les doublons des parties partagées ont en partie été enlevés par les informaticiens de l'ATILF.

7. Site : <http://cqpweb.lancs.ac.uk> ; en local à l'ATILF : <https://arcas.atilf.fr/cqpweb/>.

2.4 Citations hors bases

Il est fréquent que le lexicographe lise ou entende de façon aléatoire des énoncés qu'il note et veut utiliser comme exemples lexicographiques. Cette pratique est tout à fait normale et doit être exploitée au maximum, de tels exemples étant potentiellement d'une grande valeur lexicographique.

Pour les citations hors bases, il est essentiel de reproduire exactement ce qui a été lu ou entendu, afin de garder à ces exemples leur nature de citation. Si le lexicographe ne les a pas notés et n'en a qu'un souvenir approximatif, la retranscription qu'il donne relèvera alors de la classe suivante : les exemples fabriqués ou adaptés (section 2.5).

Les citations hors bases se font selon le même modèle que les trois autres types de citations examinés plus haut, en adaptant le référencement à la source de la citation.

- a) Citations de livres, de journaux, etc. → même référencement que pour Frantext ou les corpus journalistiques.

Attention ! Les citations extraites d'ouvrages ou documents mis à disposition sur le web – sur Google Books, notamment – entrent dans cette classe et doivent être référencées relativement à l'ouvrage dont elles proviennent. Elles ne rentrent pas dans la classe d) ci-dessous.

- b) Citations d'énoncés entendus à la radio, à la télévision ou au cinéma → suivre un des patrons de référencement ci-dessous.

[<Nom de l'émission de radio/télévision>, <Nom de la chaîne de radio/télévision>, <Date de diffusion (jour, mois et année)>]

Par exemple :

- (8) Messieurs les **censeurs** bonsoir !

[À *Armes égales*, première chaîne de l'ORTF, 13 décembre 1971]

[Réplique de <Nom du personnage> (joué(e) par <Nom de l'acteur>), dans <Nom du film>, <Année de réalisation>]

Par exemple :

- (9) - Tu sais, il est **autodidacte**. Oui, il s'est fait tout seul.

[Réplique de Anna (jouée par Juliette Binoche), dans *Mauvais sang*, 1986]

- c) Citations d'énoncés entendus dans des conversations → suivre le patron de référencement suivant :

[<Très brève caractérisation du contexte d'énonciation>, <Date (jour, mois et année)>]

Par exemple :

- (10) J'ai bien fait de prendre mon gilet, je ne me **déshabille** pas, il fait trop froid !

[Dans le train, 10 septembre 2011]

C'est au lexicographe d'évaluer le degré de précision nécessaire pour la bonne caractérisation du contexte d'énonciation, **en veillant à être le plus économe possible**. En règle générale, le lieu d'énonciation devra être spécifié, mais il peut être utile, dans certains cas, de caractériser aussi de façon approximative le locuteur et/ou le destinataire.

- d) Citation du web → suivre le patron de référencement suivant :

[<URL de la page web>, <Date de consultation de la page (jour, mois et année)>]

Par exemple :

- (11) La cérémonie officielle de rentrée des Universités de Lorraine, **se tiendra** cette année le lundi 3 octobre salle Poirel à Nancy.
[<http://www.univ-nancy2.fr/ACTU/aperçuActu.php?idActu=1658>, 3 octobre 2011]

Remarque Il serait en principe important de donner également la date de rédaction/modification de la page consultée, mais cette information est trop souvent soit non disponible, soit non fiable (mises à jour faites sans en indiquer la date). On doit se résoudre à donner une information incomplète. C'est justement une des faiblesses du web en tant que corpus et du web dans sa réalité textuelle : ce n'est pas un corpus contrôlé ou contrôlable.

e) Citation de courriels → suivre le patron de référencement suivant :

[Courriel <+ très brève caractérisation du contexte si nécessaire>, <Date Date (jour, mois et année)>]

Par exemple :

- (12) Jean-Luc s'est chargé de la **vérification** de nos exemples.
[Courriel professionnel, 26 septembre 2011]

2.5 Exemples fabriqués ou adaptés

Une section de la zone d'exemples du RLF est réservée aux exemples fabriqués ou adaptés. Par exemple, (13a) est un exemple fabriqué pour illustrer une acception du vocable FLEURIR et (13b) est un exemple adapté de l'exemple (13c), extrait de FrWac.

- (13) a. De jolis motifs verts et rouges **fleurissaient** les murs de sa chambre.
b. Aujourd'hui la jachère **fleurit** au milieu des champs puisque les agriculteurs y sèment des plantes indigènes.
c. Aujourd'hui la jachère **fleurit** au milieu des champs puisque les agriculteur y sèment des plantes indigènes.
[FRWAC, 68236516]

Comme on le voit, l'adaptation en (13b) de (13c) consiste en la simple correction d'une faute d'orthographe. On pourrait décider, s'il nous semble important de référencer l'exemple en question, de le présenter dans la zone des citations de FrWac, en indiquant simplement la présence de la faute au moyen d'un *[sic]*.

- (14) Aujourd'hui la jachère **fleurit** au milieu des champs puisque les agriculteur *[sic]* y sèment des plantes indigènes.
[FRWAC, 68236516]

L'adaptation des exemples est plus souvent utilisée pour rendre plus « élégant » ou compact un énoncé d'une certaine lourdeur. Elle sert aussi à rendre plus autonome un énoncé qui, sinon, pourrait paraître étrange ou peu clair une fois tiré de son contexte.

Dans tous les cas, il faut noter que nous n'établissons aucune distinction, quant au mode de présentation, entre exemples fabriqués – forgés de toutes pièces – et exemple adaptés : **ni les uns ni les autres ne sont référencés.**

2.6 Citations hors-norme

Une section particulière de la zone d'exemples est disponible pour des citations relevant d'un français « hors-norme ». Ces citations peuvent être de plusieurs natures, notamment :

- *greffes collocationnelles* (Polguère, 2007), qui nous semblent être des erreurs, mais dont nous pensons qu'elles ont une fréquence d'emploi significative ;
- emplois d'une lexie qui nous semblent relever d'un glissement sémantique assez bien attesté, mais dont nous ne jugeons pas la lexicalisation effective ;
- emplois d'une lexie avec un régime « anormal », mais qui méritent pour diverses raisons d'être relevés ;
- etc.

De telles citations doivent être référencées selon les patrons proposés plus haut. Elles représentent ce que l'on pourrait appeler une *zone de turbulence lexicale* dans l'utilisation de la lexie décrite. Dans tous les cas, de telles citations doivent être accompagnées d'un commentaire qui les analyse et justifie leur inclusion dans un article de RLF⁸.

3 À propos des corpus oraux

On ne dispose pas à l'heure actuelle de corpus d'envergure de productions orales en français, interrogeables au moyen d'un concordancier. C'est pourquoi nous n'envisageons pas le recours à de tels corpus dans le cadre du projet RELIEF.

Mentionnons qu'un corpus de productions orales est tout de même disponible à l'ATILF. Il s'agit du corpus du projet TCOF (Traitement de Corpus Oraux en Français)⁹, consultable en libre accès sur le site CNRTL : <http://www.cnrtl.fr/corpus/tcof/>.

Remerciements

Merci à Pascale Bernard, Sébastien Haton, Dorota Sikora et Inka Wissner pour leurs commentaires.

Bibliographie

Virginie André et Emmanuelle Canut. Mise à disposition de corpus oraux interactifs : le projet TCOF (Traitement de Corpus Oraux en Français). *Pratiques : linguistique, littérature, didactique*, 147/148:35–51, 2010.

Marco Baroni, Silvia Bernardini, Adriano Ferraresi et Eros Zanchetta. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226, 2009.

Pascale Bernard, Josette Lecomte, Jacques Dendien et Jean-Marie Pierrel. Un ensemble de ressources informatisées et intégrées pour l'étude du français : FRANTEXT, TLFi, Dictionnaires de l'Académie et logiciel Stella, présentation et apprentissage de leurs exploitations. Dans *Actes de TALN 2002*, volume 2, pages 3–36, Nancy, 24–27 juin 2002.

8. La zone d'exemples (EX) des articles RLF n'étant pas encore formalisée dans l'éditeur lexicographique DixitRLF, on se contente pour l'instant de simplement écrire les commentaires, à la ligne, immédiatement sous l'exemple concerné.

9. Cf. André et Canut (2010).

Andrew Hardy. CQPweb – combining power, flexibility and usability in a corpus analysis tool. URL <http://www.lancs.ac.uk/staff/hardiea/cqpweb-paper.pdf>. à paraître.

Alain Polguère. *Soleil insoutenable et chaleur de plomb* : le statut linguistique des greffes collocationnelles. Dans Marie-Claude L’Homme et Sylvie Vandaele, dir., *Lexicographie et terminologie : compatibilité des modèles et des méthodes*, pages 247–291. Les Presses de l’Université d’Ottawa, Ottawa, 2007.