



HAL
open science

Spiderlex et compagnie

Sandrine Ollinger, Alain Polguère, Yannick Chudy, Bruno Gaume

► **To cite this version:**

Sandrine Ollinger, Alain Polguère, Yannick Chudy, Bruno Gaume. Spiderlex et compagnie. 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 4: Démonstrations et résumés d'articles internationaux, Jun 2020, Nancy, France. pp.60-63. hal-02768517v3

HAL Id: hal-02768517

<https://hal.science/hal-02768517v3>

Submitted on 23 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Spiderlex et compagnie

Sandrine Ollinger¹ Alain Polguère² Yannick Chudy³ Bruno Gaume⁴

(1) CNRS, Université de Lorraine, ATILF, UMR 7118 Nancy, F-54063, France

(2) Université de Lorraine, CNRS, ATILF, F-54063, France

(3) La Maison de l'Initiative, Toulouse, F-31100, France

(4) CNRS, Université de Toulouse, CLLE-ERSS, UMR 5263 Toulouse, F-31058, France

Sandrine.Ollinger@atilf.fr, Alain.Polguere@univ-lorraine.fr,
ychudy@gmail.com, Bruno.Gaume@univ-tlse2.fr

RÉSUMÉ

Nous présentons l'instrument de navigation Spiderlex, mis au point pour permettre une exploration dynamique des *Réseaux Lexicaux du Français et de l'Anglais* (RL-fr et RL-en). Spiderlex est à la fois un outil de visualisation des données lexicales et un instrument de travail pour les lexicographes.

ABSTRACT

Spiderlex & Co.

We introduce Spiderlex, a lexical browser designed to allow for the dynamic exploration of the French and English Lexical Networks (fr-LN and en-LN). Spiderlex is both a tool for visualizing lexical data and for the construction of Lexical Systems by lexicographers.

MOTS-CLÉS : Système Lexicaux, ressources lexicales, visualisation de réseaux, consultation.

KEYWORDS: Lexical Systems, lexical resources, network visualization, viewer interface.

1 Introduction

Les Systèmes Lexicaux (Polguère, 2014)¹ fournissent l'ensemble des informations présentes dans les modèles lexicographiques construits selon les principes théoriques et descriptifs de la Lexicologie Explicative et Combinatoire (Mel'čuk *et al.*, 1995); ils ont la particularité d'être structurés sous forme de graphes lexicaux. Ils appartiennent à la famille des réseaux lexicaux et ont les propriétés topologiques de graphes petit-monde hiérarchiques : agrégats (angl. *clusters*) et loi de puissance des degrés (Gader *et al.*, 2014). Spiderlex, développé en collaboration avec le projet Padagraph², a été conçu pour permettre une navigation interactive sur les *Réseaux Lexicaux du Français et de l'Anglais* (RL-fr et RL-en) qui prend en compte ces caractéristiques.

Il a été pensé pour répondre aux attentes variées des utilisateurs. Les lexicographes ont besoin de visualiser quotidiennement l'impact de leur travail sur la structuration générale du réseau. Les linguistes et informaticiens intéressés par l'exploitation d'un Système Lexical sont demandeurs d'un environnement leur permettant de se familiariser avec sa structure et les informations qu'il contient. Enfin, les enseignants ont besoin de ressources pédagogiques dans le cadre d'activités d'enseignement du vocabulaire ou, au niveau universitaire, pour l'enseignement de la lexicologie/lexicographie.

1. <https://lexical-systems.atilf.fr/>

2. <https://github.com/padagraph/>

Spiderlex propose donc des fonctionnalités d'accès simples, tout en permettant un paramétrage fin pour un usage avancé.

2 Présentation générale

Le RL-fr est composé à ce jour de plus de 28 500 nœuds et 77 500 arcs. Il est entièrement construit lexicographiquement, au sein du laboratoire ATILF depuis 2011. Il a déjà atteint une grande maturité, même si sa couverture du lexique français est encore partielle. Le RL-en, quant à lui, est composé de plus de 200 000 nœuds et près d'un million d'arcs. Il a été amorcé automatiquement à partir de WordNet (Fellbaum, 1998; Gader *et al.*, 2014) et fait depuis l'objet d'un travail de correction et d'enrichissement. Pour bénéficier d'une représentation graphique comme point d'entrée dans les réseaux, sans pour autant les afficher en intégralité, nous nous sommes appuyés sur les travaux de Gaume (2008) sur la proxémie pour déterminer les meilleurs sous-graphes susceptibles de présenter l'entourage d'une unité lexicale ou d'un vocable.

Pour répondre aux besoins variés des usagers, deux modes d'explorations, l'un dit *sémantique*, l'autre *lexicale*, ont été préprogrammés. L'exploration *sémantique* est calculée à partir des poids sémantiques (indiqués dans les métamodèles) des relations entre entités lexicales, tandis que l'exploration dite *lexicale* est calculée sans pondération. Elle tient uniquement compte de la topologie du graphe. Dans les deux modes d'exploration, chaque requête de l'utilisateur donne lieu à une *balade aléatoire* (Gaume, 2004) dans le Système Lexical. Ce parcours sillonne le graphe sans restriction vis-à-vis de l'orientation des arcs, en prenant en compte les boucles éventuelles et les arcs multiples. Par défaut, la longueur de la balade est de 3 et le nombre de nœuds retournés est de 30 par entité recherchée.

L'utilisateur de Spiderlex choisit un mode d'exploration, puis interroge les entités lexicales qui l'intéressent à l'aide de leur forme ou de leur identifiant unique. Une représentation graphique du sous-graphe obtenu est affichée dans la partie centrale de l'interface. La Figure 1 montre ainsi le résultat de la consultation du vocable TORTUE, qui regroupe deux acceptions : TORTUE I [*Que mange une jeune tortue ?*], acception qui dénote un animal et TORTUE II [*C'est une vraie tortue celui-ci, il va à deux à l'heure !*], acception métaphorique, qui dénote un individu qui se comporte avec lenteur.

Dans la partie droite de l'interface, trois onglets sont disponibles : Liens lexicaux, Agrégats et Dictionnaire. Le premier permet de sélectionner quels types de liens on veut voir apparaître dans le graphe (composant sémantique, inclusion formelle, copolysémie et fonctions lexicales³). Le sous-graphe n'est pas recalculé lorsque l'on modifie ces paramètres, mais les arcs désactivés sont rendus invisibles et les nœuds qui n'étaient reliés que par eux sont grisés. Il n'est plus possible de les sélectionner. L'utilisateur peut ainsi visualiser rapidement quelles relations structurent l'espace sémantique ou lexical qu'il est en train d'explorer.

Comme le montre la Figure 2, l'onglet Agrégats permet d'observer sous forme de listes des sous-groupes de nœuds fortement liés à l'intérieur de la sélection d'entités lexicales effectuée. Ces agrégats rendent compte de sous-groupes de nœuds fortement liés topologiquement à l'intérieur de la sélection d'entités lexicales effectuée. Ils sont calculés à l'aide de l'algorithme Walktrap (Pons & Latapy, 2006), appliqué par défaut. Tout comme les paramètres de constitution de sous-graphes, la méthode de détection de communautés est modifiable dans un mode expert, où l'algorithme Infomap (Rosvall

3. Nous vous invitons à consulter la distribution publique du RL-fr (ATILF, 2020), sa documentation et sa bibliographie pour une description détaillée des différents types de liens cités ici.

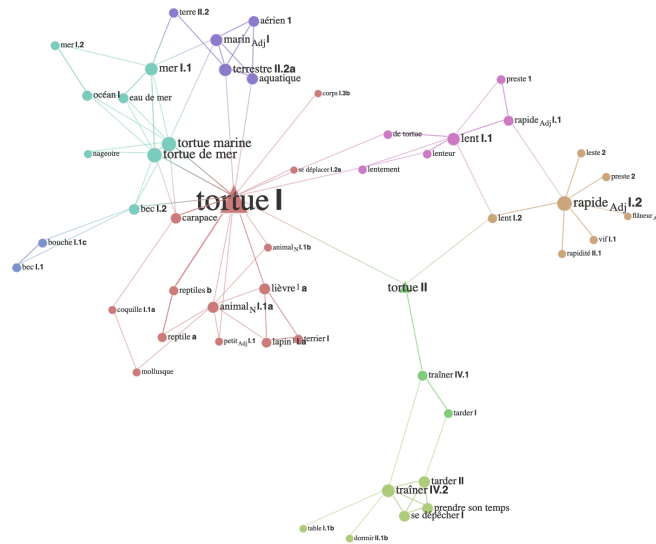


FIGURE 1 – Graphe résultant d’une exploration sémantique autour du vocable TORTUE

et al., 2009) est également disponible. On observe bien souvent que des agrégats incohérents rendent compte de descriptions insuffisantes ou de liens erronés dans la ressource.

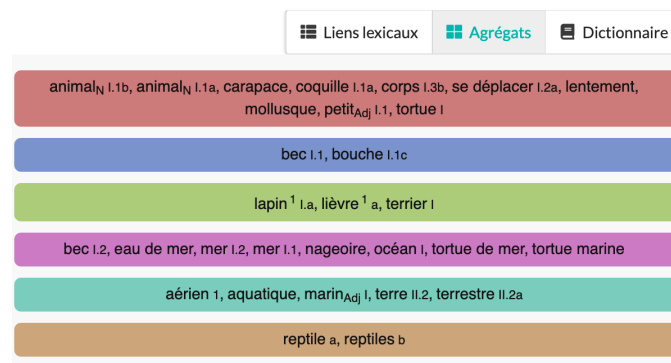


FIGURE 2 – Agrégats sémantiques dans le voisinage de TORTUE I et TORTUE II

Enfin, l’onglet *Dictionnaire*, illustré par la Figure 3 donne accès aux articles lexicographiques. Il contient l’ensemble des informations lexicales encapsulées dans chaque nœud du graphe. Dans un premier temps, les nœuds visibles sont ceux des entités lexicales recherchées. L’utilisateur peut changer d’article en sélectionnant une entité lexicale dans la vue-article ou sur la représentation graphique. Enfin, il peut lancer le calcul d’un nouveau sous-graphe en cliquant sur l’en-tête de l’article.

Spiderlex est disponible en ligne (<https://spiderlex.atilf.fr/>) et est exploité par les lexicographes des RL-fr et RL-en. Il sert aussi de source de données pour des travaux de recherche, notamment en enseignement/apprentissage du vocabulaire.

The image shows a screenshot of a French dictionary entry for the word "tortue". At the top, there are navigation tabs: "Liens lexicaux", "Agrégats", and "Dictionnaire". The word "tortue" is displayed in a search bar. Below the search bar, the entry is organized into sections: [CG] (Classification Grammaticale) showing "nom commun" and "fém"; [DF] (Définition) showing "animal" and a detailed definition: "animal (petit) dont le corps est entouré d'une carapace qui se déplace lentement terrestre ou aquatique"; [FL] (Formes Lexicales) showing synonyms ("tortue de mer", "tortue marine"), contrasts ("lièvre"), generalizations ("animal", "spéc reptiles"), and meronyms ("bec", "carapace"); and [EX] (Exemple) with a sentence from a 1975 text by Michel Tournier: "Puis comme elle s'était inquiétée des dommages causés aux plantes aquatiques par les tortues d'eau qui infestaient les bassins, Ralph entreprit en fouillant à pleins bras la vase de capturer l'une de ces bestioles." The source is cited as "Frantext TOURNIER Michel, *Les Météores*, 1975, p. 471".

FIGURE 3 – Article lexicographique de TORTUE

Références

- ATILF (2020). Réseau lexical du français (rl-fr). ORTOLANG (Open Resources and TOols for LANGuage) – www.ortolang.fr.
- FELLBAUM C. (1998). *WordNet : an electronic lexical database*. Language, Speech and Communication. MIT Press. DOI : [10.7551/mitpress/7287.001.0001](https://doi.org/10.7551/mitpress/7287.001.0001).
- GADER N., OLLINGER S. & POLGUÈRE A. (2014). One Lexicon, Two Structures : So What Gives ? In H. ORAV, C. FELLBAUM & P. VOSSE, Éd., *Seventh Global Wordnet Conference (GWC2014)*, p. 163–171, Tartu, Estonia : Global WordNet Association. HAL : [hal-00937187](https://hal.archives-ouvertes.fr/hal-00937187).
- GAUME B. (2004). Balades aléatoires dans les Petits Mondes Lexicaux. *Revue I3 - Information Interaction Intelligence*, **4**(3).
- GAUME B. (2008). Mapping the forms of meaning in small worlds. *International Journal of Intelligent Systems*, **23**(7), 848–862. DOI : [10.1002/int.20275](https://doi.org/10.1002/int.20275), HAL : [hal-01322013](https://hal.archives-ouvertes.fr/hal-01322013).
- MEL'ČUK I., CLAS A. & POLGUÈRE A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Paris/Louvain-la-Neuve : Duculot. DOI : [10.1017/S0008413100020582](https://doi.org/10.1017/S0008413100020582).
- POLGUÈRE A. (2014). From Writing Dictionaries to Weaving Lexical Networks. *International Journal of Lexicography*, **27**(4), 396–418. DOI : [10.1093/ijl/ecu017](https://doi.org/10.1093/ijl/ecu017).
- PONS P. & LATAPY M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, **10**(2), 191–218. DOI : [10.1007/11569596_31](https://doi.org/10.1007/11569596_31).
- ROSVALL M., AXELSSON D. & BERGSTROM C. T. (2009). The map equation. *The European Physical Journal Special Topics*, **178**(1), 13–23. DOI : [10.1140/epjst/e2010-01179-1](https://doi.org/10.1140/epjst/e2010-01179-1).