

A tool for linking stems and conceptual fragments to enhance word access

Núria Gala (LIF-CNRS)

Véronique Rey (SHADYC EHESS et CNRS)

Michael Zock (LIF-CNRS)

Aix – Marseille Université (France)

Electronic dictionaries

- Mainly **reader-oriented**

- Heterogeneous information:

- grammatical categories,
- meaning (definitions),
- examples of use (word's usages),
- lexically related words,
- lexical functions,
- etymology
- ...

→ What is relevant for **language production** ?

Electronic dictionaries

- Some conclusions from E-lexicography conference (Louvain, oct. 2009):

Still a lot to be done concerning:

- some hard points: word senses → usages
- **the user needs: access to new words**
- the exploitation of the electronic medium: queries, browsing, displaying information, etc.

Outline

- The speaker at the starting point
- Existing resources for French word families
- Morpho-phonological families
 - Morphological description of lexical units
 - Semantic features in a family
- Finding and producing words with Polymots
- Conclusion and further work

Starting point

- The speaker knows what s/he wants to say
- S/he knows the word...
- But s/he is unable to access it

- Tip of the tongue phenomena
- Paraphasia
- Language learning

Point of view of the language speaker

Access to words from conceptual fragments

- how do I say something 'sticky' and 'strong' in English?

Access to words from formal relationships

- what's the word for a 'piece of clothe' or a 'band on the arm'?

Writing a word with the appropriate orthograph

- do 'time' or 'weather' take a 'p' in French?

Aim of our work

- Capitalize on the bidirectionnal links between
 - Semantics → conceptual fragments
 - Morpho-phonology → stems
- Present a resource for **French words grouped into morpho-phonological families**
- Propose such a resource
 - for vocabulary and orthography learning
 - from a language producer's point of view
 - to be used for education and by speech therapists

Existing resources

- Few resources to help the learner to **acquire new vocabulary** and/or to **master spelling** on the basis of 'families'
- Different concepts for 'word family' depending on the way lexical units are considered:
 - (a) Etymological families (evolution)
 - (b) Analogical families (synonymy)
 - (c) Thematical families (domain)

Etymological families

- Diachrony : **words evolution** in time
- Words sharing a 'canonical form' or a 'lexical root' generally at the beginning of the creation of other words in the family

- Ex. Synapse

<http://www.synapse-fr.com>

</produits/Famille.htm>

The screenshot shows the 'Fenêtre globale d'accès aux dictionnaires' window. The search term 'parent' is entered in the search bar. The left sidebar lists various linguistic tools, with 'Familles de mots' selected. The main area displays the search results for 'parent', including a list of related words and a detailed definition of 'parent' and 'parentales'.

parent

Tête de la famille du mot : parent

type grammatical : substantif

apparentage [nom]
apparenté [adjectif]
apparemment [nom]
apparenter [verbe]
parentage [nom]
parentaille [nom]
parental [adjectif]
parentales [nom]
parentaux [adjectif]
parenté [nom, au sens 3]
parenté [nom, aux sens 1,4]
parenté [nom, au sens 2]
parenté [nom, au sens 5]
parentèle [nom]

parentales
= parentales
Nom féminin pluriel
dans l'Antiquité, à Rome, fêtes annuelles célébrant les morts.

Analogical families

- Similarity, **close meaning**, same referent in the world
- Ex. Centre Collégial de Développement de Matériel Didactique du Québec

http://www.ccdmd.qc.ca/fr/jeux_pedagogiques/?id=1089&action=animer

FAMILLES

Module **D**

Série 1 de 4

Regroupez les mots en trois familles.

entraver

paralyser

languir

déjouer

ballotter

dérouter

interdire

agiter

secouer

souffrir

empêcher

touiller

pâtir

brasser

supporter

remuer

endurer

tolérer

Quitter

Thematical families

- **Term associations** made by humans
(broom → household, cleaning, house...)
- Lexical networks being used by machines
- Ex. JeuxdeMots
(Lafourcade, 2007)

<http://www.lirmm.fr/jeuxdemots/generateGames.php>



POLYMOTS

- A resource for learning words on the basis of morpho-phonological families
- A family is a group of lexical units sharing:
 - **Formal analogies**: common stems
 - alternations are possible
 - **Semantic continuum for users**: similar conceptual ideas for the speaker
 - the degree of semantical cohesion in a family may vary

POLYMOTS

- a **phonological structure**:
 - *bras, brassard, bracelet, embrasser...* /bRa/
 - *temps, temporel, température...* /t@/
 - *preuve, prouver, approbation...* /prØv/ ~ /pruv/
- a **semantic coherence** for users:
 - *vallée, avaler, avalanche...* → going downhill
 - *accident, suicide, acide...* → death, danger
 - *glu, agglutiner, gluant...* → sticky, strong, together

POLYMOTS

- The process of word construction implies morpho-phonological transformations: vocalic and consonantic alternations (Kiparsky, 1982)
- Keeping the phonological form of a lexical unit as a memory help: minimal listing or stem-only hypothesis (Taft, 1981)

POLYMOTS

- Recognizing a link between two objects can lead to create a word on the basis of formal and semantic analogies keeping
 - the stem (ground: '*terre*'; moon: '*lune*')
 - one or some ideas
 - surface: '*terrasse*'
 - moon-shaped, roundness: '*lunnettes*'

Methodology (1)

- Manual **global segmentation** of a list of 20,000 words
 - stems identified afterwards, in synchrony
- Multiple occurrences
 - a stem being a lexical unit (*chaise, écran, falaise*)
 - or being shared by a list of units (*bouleau, boulette, boulier...; terre, enterrer, terrasse...*)

Productivity

- 20,000 words, 2,004 stems = families
- The more general the stem's meaning, the larger the family

Number of words	Number of families	
1	90	<i>autel, chaise, mot, paupière ...</i>
2 to 3	312	<i>acier, alcool, fée, éternu, souris ...</i>
4 to 5	430	<i>abeille, caprice, poisson ...</i>
6 to 7	322	<i>alphabet, lot, nord, oeil ...</i>
8 to 9	185	<i>ange, canon, drame, fisc, vache ...</i>
10 to 20	441	<i>ample, fer, figure, monnaie ...</i>
> 20	224	<i>acte/ag, forme, mode, port ...</i>

Methodology (2)

- Semi-automatic **acquisition of conceptual fragments** from available lexical and encyclopaedic resources (Gala & Rey, 2009):
 - definitions from Wiktionnaire,
 - introductory paragraph in Wikipedia
- Grouping, filtering and weighting conceptual fragments
- Construction of semantic vectors

Examples

thematic links

synonyms

common stems / common semantic units / hyperonyms

Embrasser

[serrer 1] [contenir 0.66] [saisir 0.66]
[bras 0.58] [attacher 0.44] [entourer
0.44] [étendre 0.32] [regard 0.32]
[adopter 0.29] [baiser 0.25] [englober
0.16] [êtreindre 0.15] [engager 0.13]

Avaler

[descendre 1] [abaisser 0.48]
[accepter 0.38] [gosier 0.32]
[manger 0.32] [couper 0.19]
[mâcher 0.16] [supporter 0.09] ...

Vache

[femelle 1] [mammifère 0.58]
[domestique 0.54] [ruminer 0.50]
[porteur 0.45] [espèce 0.43]
[corner 0.41] [front 0.37]
[appartenir 0.32] [adulte 0.31]
[manoeuvrer 0.31] [peau 0.31]
[récipient 0.31] ...

Alarme

[signal 1] [ennemi 0.75] [arme 0.71]
[approcher 0.69] [prévenir 0.43]
[dispositif 0.40] [surveillance 0.38]...



POLYMOTS

Recherche



Recherche alphabétique



Recherche par sens



Recherche par type



Recherche simple

Ce formulaire vous permet d'obtenir, pour un mot tapé, la liste des mots de sa famille morpho-phonologique. Vous pouvez également obtenir la liste des mots dérivés qui contiennent la sous chaîne tapée. Cliquez sur un mot dans la liste pour afficher ses caractéristiques.

Rechercher un mot



arme

Lancer

Résultats

arme
 alarme
 alarmer
 alarmisme
 alarmiste
 armable
 armateur
 armature
 arme
 armement
 armer
 arneur
 armeuse
 armistice
 armoire
 armoiries
 armorier
 armure
 armurerie
 armurier

26 résultats

Fiche détaillée de "alarmisme"

Mot base : arme

Nombre de mots dérivés contenant le mot base : 25

Type : transparent

Productivité du mot base :

1.25 %

Sélectionnez un affixe ou un sens pour afficher tous les mots dérivés ayant respectivement le même affixe ou le même sens.

Sens :

alarmiste
 esprit
 état
 frayeur
 répandre
 systématiquement
 tendance

Affixes :

al
 isme

LIF-TALEP, Aix Marseille Universités, 2007-2010

[Accueil](#) [A propos](#) [Contacts](#)


Cette création est mise à disposition sous un [contrat Creative Commons](#)



POLYMOTS

[Recherche](#)[Recherche alphabétique](#)[Recherche par sens](#)[Recherche par type](#)

Recherche par sens

Entrez le ou les sens dont vous souhaitez la liste des mots ayant ce ou ces sens.
Le résultat sera composé des mots dérivés associés à au moins un des sens recherchés.
Exemple, rentrez : militaire,guerre,instrument si vous recherchez 'arme'.

Recherche par sens (si plusieurs sens, séparer par une ",") : Sélectionner un sens

Résultats

alarme
alarmer
alarmiste
alerter
avertisseur

5 résultats

Veillez sélectionner un élément dans la liste.

LIF-TALEP, Aix Marseille Universités, 2007-2010

[Accueil](#) [A propos](#) [Contacts](#)



Cette création est mise à disposition
sous un [contrat Creative Commons](#)

Conclusions

- A resource for **lexical access** on the basis of morphological and semantic grouping
- A tool for helping to learn vocabulary and spelling via word families
- A resource offering new functionalities of navigation: words grouped into clusters

Future work

- Exporting data to a standard format (TEI)¹
- Polymots online (fall 2010)
- Improve coverage
- Exploring portability to other languages (i.e. Romance languages)

1) Many thanks to L. Romary !



POLYMOTS

Thanks

Thankful

Thankfulness

[appreciation, grateful, gratitude, expression, glad]