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

**Nuria Gala[1], Véronique Rey[2], Michael Zock[3]**
[1,3] LIF- CNRS, UMR 6166, F-13288 Marseille
[2] SHADYC CNRS & EHESS, UMR 8562, F-13002 Marseille
E-mail: nuria.gala@lif.univ-mrs.fr, veronique.rey-lafay@univmed.fr, michael.zock@lif.univ-mrs.fr

**Abstract**

Electronic dictionaries offer many possibilities unavailable in paper dictionaries to view, display or access information. However, even these resources fall short when it comes to access words sharing semantic features and certain aspects of form: few applications offer the possibility to access a word via a morphologically or semantically related word. In this paper, we present such an application, POLYMOTS, a lexical database for contemporary French containing 20.000 words grouped in 2.000 families. The purpose of this resource is to group words into families on the basis of shared morpho-phonological and semantic information. Words with a common stem form a family; words in a family also share a set of common conceptual fragments (in some families there is a continuity of meaning, in others meaning is distributed). With this approach, we capitalize on the bidirectional link between semantics and morpho-phonology : the user can thus access words not only on the basis of ideas, but also on the basis of formal characteristics of the word, i.e. its morphological features. The resulting lexical database should help people learn French vocabulary and assist them to find words they are looking for, going thus beyond other existing lexical resources.

## 1. Introduction

Modern dictionaries of French tend to be electronic reincarnations of existing printed dictionaries (Petit Robert, Larousse, Hachette, etc.).In such resources, entries (headwords) are presented as encapsulated units containing a rich set of heterogeneous information (definition, part of speech tags, examples of usage, etymology, etc.). While in principle task independent, in practice such resources serve mainly the 'reader': given some word, he may look up its meaning, i.e. definition, grammatical information, spelling, etc. Of course, to some extent he can also get information relevant for language production: the word's usage (social practice), lexically related words (synonyms, antonyms), and in the case of WordNet, hypernyms, meronyms, etc. The possibilities of electronic dictionaries are enormous, be it with regard to layouts, presentation formats (views), or navigation (hyperlinks, etc.). Unfortunately, only a fraction of this is used. The lacking features concern above all the language producer. Yet, shifting focus from the receptive aspects of language to the productive side requires certain changes and add-ons during the dictionary building process. These changes may concern content, organization and indexing. Also, one aspect where nearly all dictionaries fall short is when it comes to access morpho-semantically *related words*, i.e. words sharing semantic features and formal aspects. For example, in current dictionaries one cannot access *derivation* on the basis of *river*, although both words have a common stem 'riv' and share the idea of 'boundary' and 'direction', a 'derivation' being a deviation from its initial 'direction'.

While most dictionaries do a quite decent job in the case of decoding, i.e. comprehension, they are much less successful when it comes to expression, language production. One of the problems lies with the input. In what terms specify the query or conceptual input (concepts, primitives, words) in order to get the corresponding lexical form? Another problem resides in the fact that inputs tend to vary[1]and to lack specificity. Conceptual input is often underspecified, the word's meaning or definition being only partially available.

A special case, though not all that rare, is the *tip-of-the-tongue state* (TOT), where the person knows the meaning but fails to access the corresponding form (Brown & McNeill, 1966). In such cases, few dictionaries if any are of real help (Zock & Schwab 2008). The reason for this lies probably in the fact that most dictionaries have been built from the reader's perspective. Nevertheless there have been attempts (in particular for English) to build navigational tools serving also the language producer. For example, *thesauri* (Péchoin, 1992; Rogets, 1852), Longman's *Language Activator* (1993), *analogical dictionaries* (Boissière 1862; Robert et al. 1993; Niobey et al. 2007) and, of course, WordNet (Miller, 1990).

The goal of this paper is to present POLYMOTS, a lexical database for French revealing and capitalizing on the bidirectional links between semantics and morpho-phonology. In the next section we present some existing tools based on the notion of *word families*. In section 3, we sketch the main features of our resource: morpho-phonological analogies and semantic characterization of the lexical items, the latter entailing the notion of semantic continuity. Before concluding, we will consider the use of such a database for language production.

[1] We do not always have the same set of concepts or conceptual fragments in mind when we initiate search. Suppose you were thinking of a 'dog'. In some cases the notion 'cute' is part of the message, whereas in others we just think of a 'mad animal'.

## 2. Word families: an overview

There are quite few resources presenting words in terms of families (see below), one of their aim being to help the learner to acquire new vocabulary. Lexical items can be grouped into different *families*, all depends on the view-point: evolutionary (etymology), semantic (analogical, thematic, i.e. domain) etc. We offer here a another viewpoint, *morpho-phonological families*. This entails a new way of considering words in a language.

### 2.1 Families based on etymology

Traditionally, the notion of 'word family' is concerned with etymology. In this approach language is studied diachronically, that is, the focus is on the words' evolution in time. Lexical items in an etymological family share a lexical root called 'canonical form' or 'headword' which is considered to be at the beginning of the creation of what is called a derived lexical unit. The following words '*change*s', '*change*d', '*change*ful', 'ever-*chang*ing', 'inter*change*ably', 'inter*change*ability', etc., are all derived terms, sharing the lexical root 'change'. When studying a word's etymology, one does not take into account inflectional characteristics. In the example given, 'changes' and 'changed' are morphological variations of the same lexeme (their part-of-speech is identical: 'change' and 'changes' are verbs or nouns, 'change' and 'changed' are verbs). Etymological families are thus concerned with word formation, i.e. the way new words are created: *derivation* (use of affixes as in 'inter-change-ably') and *compounding* (combining different words as 'ever' and 'change' into a new word like 'everchanging').

A well-known resource for etymological families in French is the dictionary built by Synapse[2], which groups words according to meanings (proximity) and form (morphological similarities). For instance, the term 'parent' is linked with 'ap*parenté*' (related), '*parent*èle' (relatives), '*parent*alité' (parenthood) etc. The derived forms are displayed by meanings in case of a polysemic headword. However, the lexical units being tagged only in terms of part-of-speech, their underlying semantics is not made explicit.

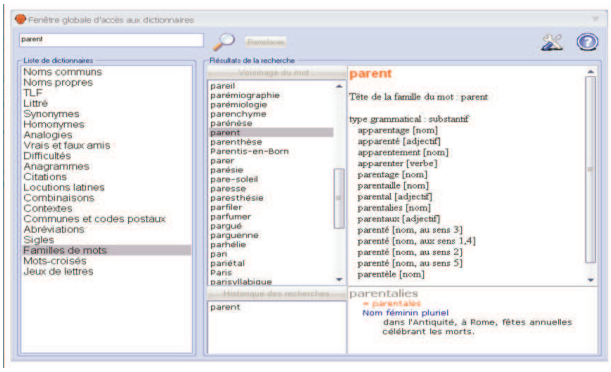


Figure 1. The entry *'parent'* in the Synapse dictionary.

[2] http://www.synapse-fr.com/produits/Famille.htm

### 2.2 Families based on meaning

Another way to group words is by similarity, for example synonymy, analogy, etc. Synonymy deals with words having basically the same referent in the world. This way of grouping is based exclusively on semantic considerations, disregarding entirely formal aspects. Whatever the words' morphological structure, words are grouped according to their semantic similarity: 'change' is semantically close to 'alter', 'transform', 'modify', etc..

The CCDMDQC (Centre Collégial de Développement de Matériel Didactique du Québec) [3] offers a resource for French based on analogies. By playing a game, the learner drags word tags and aligns them according to semantic criteria. For example, the following terms, 'cher' (expensive), 'coûteux' (costly) and 'precieux' (precious) may all be grouped in a single family. The same holds true for 'bon' (good), 'appétissant' (tasty) and 'délicieux' (delicious).

### 2.3 Families based on domain

While the just mentioned resources have been developed to help learners acquire or expand French vocabulary (be it a first or second language), *thematic families* are based on typical associations of terms (racket, tennis, net) and may also be used by machines, i.e. natural language applications. Thesauri are typical representatives of this group. Yet, words can also be grouped according to their semantic relations: *superordinate*/*subordinate* ('animal' and 'dog'), *part*/*whole* ('roof' and 'house'), or Mel'cuk's *lexical functions* (i.e. 'magn' as in 'fever' and 'high'). A lexical resource illustrating this kind of approach for French is JeuxdeMots[4] (Lafourcade, 2007), built collaboratively via a game. The goal is to find out what people typically associate with each other and to build then the corresponding lexical-semantic network. What is a typical association is established empirically: given some input from the system (term and link) the user produces the associated word (second term), answering this way the question, what term x is related with y. Once the network is built, terms can be found by entering the network (via some input) and by following the links until the target word is found.

## 3. Morpho-phonological families

Having seen a sample of applications for French based on the notion of 'word family', we present here below POLYMOTS, a database describing the lexical organization of contemporary French in terms of semantic and phonological information (Kiparsky 1982; Troubetzkoy 1964). The application deals exclusively with contemporary French and is based on such notions as formal analogy and semantic continuity. While recognizing a common stem in different words is quite straightforward, seizing the notion of *semantic conti-*

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

[4] http://www.jeuxdemots.org

*nuum* is more difficult, as it is less objective. Indeed, the common idea in 'bras' (arm), '*brac*elet', '*bras*sard' (armband') and 'em*bras*ser' (kiss, embrace) is much more obvious than in the following list: 'val' (valley), 'av*al*anche' (flood), 'av*al*er' (swallow). Even if the commonality is hard to perceive immediately, there is one, the idea of *going downhill* being present in all the members of the list. Hence, all these words can be ascribed to belong to the '*go downhill*'-family. Likewise, the idea of 'ride' (wrinkle) shows up in words like '*ride*au' (curtain) and '*ride*lle' (slatted side). We consider all these words as members of the same family, as they share morpho-phonological forms ('tort' /toR/, 'ride' /Rid/) as well as semantic features (conceptual fragments).

### 3.1 Morphological description of lexical units

As mentioned, words with a common stem form a family. A stem can appear alone (case of certain lexemes which are ordinary words having a meaning) or as part of a word. This principle allows us to distinguish between *transparent* stems (about 75%) and *opaque* stems (about 25% of the database). For example, 'fil' (thread) is the common, transparent stem in 'dé*fil*é' (parade), 'pro*fil*' (profile), '*fil*iation' (parentage), '*file*' (queue); 'cid' is not a lemma anymore in modern French, but it is the common opaque stem in 'ac*cid*ent' (accident), 'sui*cid*e' (to commit suicide), 'dé*cid*er' (to decide), and 'a*cid*e' (acid), etc.

In terms of productivity, the number of elements of a family depends on the stem's meaning: the more general the meaning, the larger the family.

While some families have only one or two members (époque-epoch; abri-shelter; *abri*ter-to shelter), others have 70-100 lexical items, 'act' and 'fact/fit/fait' being examples in case: 'act' in '*act*ivité, ré*act*ion, *act*uel, *act*eur, con*tact*er', etc. (activity, reaction, actual, actor, to contact); 'fact/fit/fait'[5] in con*fit*ure, dé*fait*e, édi*fic*e, for*fait*', etc. (jam, defeat, building, daily pass).

### 3.2 Semantic features of family members

POLYMOTS represents words as a vector of semantic units (conceptual fragments) obtained automatically from structured corpora (Gala & Rey 2009). For example, 'cow' is described by a vector containing, among other, *female, mammal, domestic, ruminate, milk*; 'alarm' is described by *signal, ennemy, weapon, device, monitor*, etc. The features have weights, which shows their relative importance with regard to the headword.

While phonologic grouping of words is certainly quite useful, it raises nevertheless several questions concerning the semantic organization of the lexicon. For example, what is common between the following pairs of French words: 'arme' - 'al*arme*' (*weapon-alarm*), 'ré*ac*-tion' - '*act*eur (*reaction - actor*), or 'ac*cid*ent' - 'a*cid*e' (*accident - acid*)? While in some families there is a continuity of meaning (words sharing a significant number of semantic features[6]), in others meaning is distributed. The semantic features of the common stem are shared by the members of the family: 'val' (glen) includes the features *geographic area* and *going downhill*, and at least one of them is also present in 'vallée' (valley) and 'avaler' (to swallow).

## 4. Ways to access words using Polymots

Polymots offers different functionalities to access words and word families: keywords, lists, productivity and meaning.

### 4.1 Queries by keyword or substring

By typing a word, the resource provides the list of all the lexical units of the family; by selecting one of them in the list, the application would show the result of morphological (list of affixes) and semantic analysis (list of semantic units describing the selected lexical unit). Figure 2 illustrates the result for a query with the keyword 'acteur' (actor):

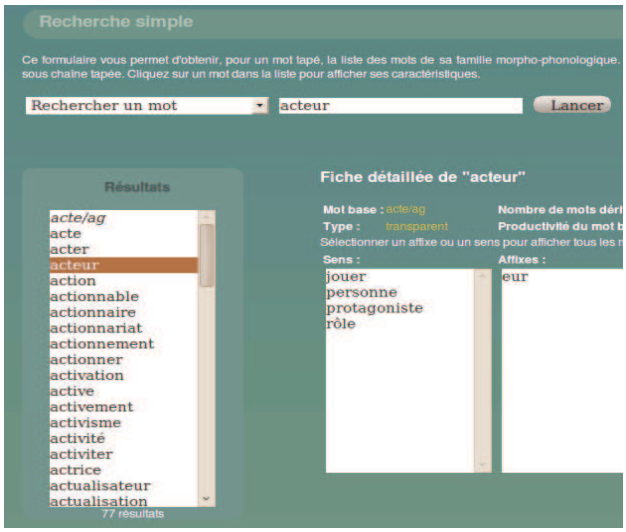


Figure 2. Result of a query for the keyword 'acteur'.

Family members are shown in the left window, with the stem in italics at the beginning of the list. Once a word is selected, POLYMOTS displays information in the other two windows. For example, it will provide detailed semantic and morphological information concerning the item under scrutiny: in the middle, conceptual fragments (to play, person, main character, role), and in the right column, suffixes (-eur).

Typing a substring, i.e. 'arg', would yield a list of all the lexical items containing it, regardless of family membership: 'charge', 'ép*arg*ne', 'h*arg*ne', 'l*arge*', 'm*arge*', etc. (load, savings, spite, wide, margin, etc.). Selection of a stem would yield an exhaustive list of family members (selecting 'marge' would trigger a new text window

[5] Some stems have phonological alternations. (allomorphisms).

[6] This is the case of "globe/earth" ('terre'), "territory" and "terrace" which share the notion of *area* and *surface*.

containing all the words of that family: '*émarge*r', '*marge*lle', '*marg*inal', etc. (to sign, edge, marginal, etc.); selecting a word would result in revealing in a new window the semantic and morphological information concerning the selected item.

### 4.2 Queries by lists

POLYMOTS offers different kinds of queries, producing output in listform. If the user keyed in or selected a particular letter of the alphabet, he would get all the words starting with this letter. By selecting a stem, he would get all the stems corresponding to the selected type. Finally, if the query were by affix (figure 3: '-ette'), the user would get all the items having a particular prefix or suffix.


Figure 3. Result of a query by the suffix '-ette'.

Note that POLYMOTS also provides information concerning the number of words matching a selected item. In the example here above, 610 suffixes, 205 words having the suffix '–ette'.

### 4.3 Queries by productivity

POLYMOTS can also provide information concerning the productivity of the various families. For example, it is possible to search for families containing a specific number of words or affixes appearing a specified number of times (see figure 4).


Figure 4. Results for families with 60 to 90 lexical items.

### 4.4 Queries by meaning

The set of semantic features known by POLYMOTS allows for an original way to access words. By typing words as if they were conceptual fragments, the user can access all the words in the database described by them. A family can be searched via semantic features. For example, *wheather* and *forecast* would attract most of the words of the 'see-family' (vue/voi/vis) containing the prefix 'pré'.


Figure 5. Result of a semantic-based query.

Since semantic information has been added after the construction of the phonological families and not right at the beginning via an a priori set of categories (animate, human, etc.), words can now be accessed on the basis of conceptual fragments rather than via definitions as is usually the case in dictionaries[7]. This is somehow akin to lexical access in Chinese where words are indexed in terms of radicals and strokes. Word access can also be performed via morpho-phonological information, which, like syllables, are a special kind of association.

As conceptual fragments have been obtained from three different sources (Hachette dictionnary, Wiktionary and Wikipedia), the acquisition being semi-automatic, they describe a word from a wider perspective then traditional dictionaries in their definitions. This is an advantage. Unfortunately, this method has also its downside: in case of proprer nouns, certain thematic conceptual fragments might be undesirable (homonyms). Yet this can be avoided by filtering named-entities. We are currently doing some work going in this direction.

A semantic vector might thus include synonyms, hyperonyms and also thematic links. This allows us to access the word 'plombier' (plumber) via any of the following terms: 'ouvrier', 'canalisation', 'eau' and 'polonais' (worker, pipe, water, polish), all of them being syntagmatically or thematically linked to the target item, 'plombier'. Please note, that the morphem 'plomb' (lead)

[7] There have been efforts to allow word access on the basis of conceptual fragments (bag of words) extracted from definitions (Dutoit & Nuges, 2002) among others.

does not appear in this list, yet it is the transparent stem of the family, that is, 'plomb' is a real French word that could have been part of a definition in an existing lexicon. This example clearly shows that lexical resources do not focus on the construction of words, but rather on the semantic and thematic aspects of a lexical item. In this sense POLYMOTS is original: words can be accessed via semantics and morphology.

Integrating the notion of word families into a lexical database allows to reveal the fact that a given meaning can participate in various construction processes. For example, some words are the result of an analogy: 'fil' (thread: long, continuous) and 'dé*fil*é' (stroll down the street: long, continuous). Others are the result of a description: 'maintenir' (to keep) means literaly 'tenir à la main' (to hold by the hand), 'chèvrefeuille' (honeysuckle) refers to the leaves (feuilles) goats (chèvres) like to eat. Let's now see how all this relates to the mental lexicon.

## 5. Language production and organization of the mental lexicon

Despite the enormous amount work devoted to the mental lexicon (Libben and Jarema, 2007 ; Marquer, 2005 ; Bonin, 2004, Aitchison, 2003 ; Taft, 1991) many points are still unclear. A recurring topic though has been the *relationship* between the items stored. While there is a large consensus that the mental lexicon is a complex multidimensional network (Collins and Quillian, 1969), items being connected in multiple ways, it is still not entirely clear yet what the nature of the nodes and their connections are. Are the nodes single words (lemmata), smaller or bigger entities (primitives, compounds, idioms), or can they be both? The whole issue is somehow related to the very nature of words, their representation and storage. Indeed one may wonder what is actually stored: whole words (lexemes), components (stem + inflections), or also larger expressions? There is also the question whether words and their associated information (meaning, grammatical information) are stored together, locally, encapsulated like in paper dictionaries, or whether the different parts (the word's meaning, form and sound) are distributed across various layers in the network as suggested by researchers working in the spreading activation or connectionist framework (Dell et al. 1999 ; Levelt et al. 1999) .

Basically there are four questions : (a) what is stored and retrieved ? (b) what needs to be computed (inflections) ? (c) what is available at the moment of a query ? (d) how can we bridge the gap between available information and is the desired target word? By taking a look at the empirical work cited in the literature [8]it becomes clear that our lexicon has an internal structure, items being connected in various ways. WordNet is the best known resource taking this fact into account (Miller, 1990).

[8] (Aitchison, 2003 : 126-136 ; Handke, 1994 : 51-61; Harley, 2004 : 160-62 ; 240-44 ).

Not all words are lexical entries though. Hence the inflected plural form of *cow* is not a separate entry, neither are *walked* or *smarter*. On the other hand, *irregular forms* (ate, went, etc.) seem to be listed separately, so are *derivations* (nation, nationalize, nationalization). Concerning representation and storage there have been various proposals, ranging from the *full-listing hypothesis*, all words being stored fully assembled (Butterworth, 1983), to *minimal listing-* or *stem-only* hypothesis (Taft, 1981). There is also an intermediate position, the *partial-listing hypothesis* (Sandra, 1990), suggesting to list fully only common and frequent words. This makes sense, as listing all inflections is very uneco-nomical given the fact that most of them can be derived via a simple rule.

While the issue is still not yet settled, serious doubts have been raised concerning the full-listing hypothesis. Hankamer (1989) argues, that in the case of agglutinative languages (Finnish, Turkish, Hungarian) where words are formed via morpheme concatenation, words can become extremely long, hence challenging our memory (storage and access) if stored in their fully assembled form. Miller (1978) draws our attention to the fact, that even in non-agglutinative languages like English, there are phenomena speaking against the full listing hypothesis for all words. The example he gives are *number names* which are known for their productivity. Given their unlimited number makes storage in the mental lexicon impossible.

Obviously, the issue here at stake is to find a good compromise between storage and access. With regard to POLYMOTS the last two questions mentionned here above are relevant : (c) what is available at the onset of a query ? (d) how can we bridge the gap between the information available at a given moment and the target word ?

## 6. Discussion

While we cannot answer currently the question whether people really activate all the morphemes described in our work, we do believe though that our application is a good testbed to check this empirically. POLYMOTS will also allow us to check whether this kind of information helps bridging the gap between the known (input) and the unknown (target word). Inspecting logfiles and using verbal protocols may allow us to find out what is on the authors' mind when they are looking for a word without being able to find it, that is, when they feel the need to resort to a dictionary.

Wordfinding problems have been studied extensively and are known either under the headings of the TOT-problem (Brown and McNeill, 1966), or in its acute version, as the Wernicke aphasia. While the first can hit anyone, occurring only occasionally, the second is clinical, occurring regularly. People struck by this aphasia tend to make up new words, be overly verbose and produce improper word substitutions, known as paraphasia ('telephone' instead of 'television'). Just like people being in the TOT state, people experiencing

paraphasia know some information concerning the target word: aspects of sound, meaning or usage. For example, they may recall the object's function (i.e., "it serves to cut"), the first syllable (it begins by "pa") or the initial phoneme ("it begins by /k/"). In both cases (TOT and paraphasia) people are able to recognize the target word if presented in a list. Until today, there seems to be no lexical database based on the notion of word families allowing to address this problem. Yet, no doubt, such an application would be very useful.

## 7. Conclusion

This paper presents a resource for lexical access on the basis of morphological (families of words sharing a phonological stem) and semantic grouping. The goal of this kind of work is twofold. On one hand, we want to help students to learn French vocabulary and spelling via morpho-phonological families. On the other hand, we want to explore new functionalities of navigation by grouping words into clusters in order to speed up the search process. This goes clearly beyond other existing analogical resources.

The approach taken by POLYMOTS is innovative for at least two reasons. First, rather than stressing the grammatical features of morpho-phonology, we capitalize on the bidirectional link between semantics and morpho-phonology. Second, we allow the user to access words not only on the basis of ideas, but also on the basis of formal characteristics, the lexeme's morphological features. Unlike other morphological databases, POLYMOTS uses form-related information not only to reveal the construction of words, i.e. the way how they are built, but also how to find them.

## 8. Acknowledgements

We would like to thank L. Tichit for his precious help when implementing POLYMOTS, as well as the three anonymous reviewers for their constructive comments.

## 9. References

Aitchison, J. (2003). *Words in the Mind: an Introduction to the Mental Lexicon*. Oxford, Blackwell.

Boissière, P. (1862) *Dictionnaire analogique de la langue française : répertoire complet des mots par les idées et des idées par les mots*. Paris: Aug. Boyer.

Bonin, P. (2004). *Mental Lexicon: Some Words to Talk about Words*. Nova Science Publishers

Brown, R & McNeill, D. (1966). The tip of the tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5, 325-337

Burke, D.M., D.G. MacKay, J.S. Worthley & E. Wade (1991) On the Tip of the Tongue: What Causes Word Finding Failures in Young and Older Adults?, *Journal of Memory and Language* 30, 542-579.

Butterworth, B. (1983). Lexical Representation. In, Butterworth, B. (Ed.) *Language Production*, vol.2, London, Academic Press.

Collins, A. & Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior* 8 (2): 240–248

Dell G., Chang, F. & Z. Griffin (1999), Connectionist Models of Language Production: Lexical Access and Grammatical Encoding, *Cognitive Science*, 23/4, pp. 517-542.

Gala N. & Rey V. (2009) Acquiring semantics from structured corpora to enrich an existing lexicon. In *E-lexicography in the 21st century: new applications, new challenges*. Louvain-la-Neuve, Belgium.

Handke, J. (1994). *The structure of the lexicon: human vs. machine*, Mouton.

Hankamer, J. (1989). Morphological parsing and the lexicon. In W. Marslen-Wilson (Ed.), Lexical representation and process. Cambridge, MA: The MIT Press.

Harley, T. (2004). The psychology of language. *From Data to Theory. Psychology Press*, Taylor & Francis. New York

Kiparsky P. (1982) From cyclic Phonology to lexical Phonology. *The structure of Phonological Representations* (1). V. H and S. N. New York, Foris Dordrecht: 131-175.

Lafourcade, M. (2007) Making people play for Lexical Acquisition. In Proc. SNLP 2007, 7th *Symposium on Natural Language Processing*. Pattaya, Thailande

Laureys, T., De Pauw, G, Van Hamme, H, Daelemans, W. & Van Compernolle, D. (2004) Evaluation and adaptation of the Celex Dutch morphological database. Tilburg University. (Netherlands).

Levelt W., Roelofs A. & A. Meyer. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-75.

Libben, G. & Jarema, G. (2007). *The mental lexicon.* Elsevier

Longman Language Activator (1993) *The world's first production dictionary*, Longman, London

Marquer, P. (2005). *L'organisation du lexique mental.* L'Harmattan.

Miller, G. A. (1978) Semantic relations among words. In M. Halle, J. Bresnan, & G. A. Miller (eds.), *Linguistic Theory and Psychological Reality*. Cambridge, Mass.: MIT Press.

Miller, G.A., (1990). WordNet: An On-Line Lexical Database. International Journal of Lexicography, 3(4).

Niobey G., De Galiana T., Jouannon G. , Lagane R. (2007) Dictionnaire Analogique. Paris : Larousse.

Péchoin, D. (1992) (ed.) *Thésaurus Des idées aux mots, des mots aux idées*, Larousse, Paris

Robert, P., Rey, A. & Rey-Debove, J. (1993) *Dictionnaire alphabétique et analogique de la Langue Française*. Paris : Le Robert.

Roget, P. (1852) *Thesaurus of English Words and Phrases*, Longman, London.

Sandra D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *Quarterly Journal of Experimental Psychology*, 42A, 529-567

Taft, M. (1981). Prefix stripping revisited. *Journal of Verbal Learning and Verbal Behavior*, Volume 20, Issue 3, June 1981, Pages 289-297

Troubetzkoy, N. (1964) *Principes de Phonologie*. Translated from german by J. Catineau. Paris, Klincksieck, 1964[1939].

Zock, M. & Schwab, D. (2008) Lexical access based on underspecified input. In M. Zock & C. Huang (Eds.) *Proceedings of COGALEX workshop, COLING* (9-17).