Towards a resource based on *users' knowledge* to overcome the *Tip-of-the-Tongue* problem

Michael Zock
AMU (LIF, CNRS)
163 Avenue de Luminy
13288 Marseille / France
`michael.zock@lif.univ-mrs.fr`

Chris Biemann
Language Technology Group
Vogt-Kölln-Straße 30
22527 Hamburg / Germany
`biemann@uni-hamburg.de`

Abstract

Language production is largely a matter of words which, in the case of access problems, can be searched for in an external resource (lexicon, thesaurus). In this kind of dialogue the user provides the momentarily available knowledge concerning the target and the system responds with the best guess(es) it can make given this input. As tip-of-the-tongue (ToT)-studies have shown, people always have some knowledge concerning the target (meaning fragments, number of syllables, ...) even if its complete form is eluding them. We will show here how to tap on this knowledge to build a resource likely to help authors (speakers/writers) to overcome the ToT-problem. Yet, before doing so we need a better understanding of the various kinds of knowledge people have when looking for a word. To this end, we asked crowdworkers to provide some cues to describe a given target and to specify then how each one of them relates to the target, in the hope that this could help others to find the elusive word. Next, we checked how well a given search strategy worked when being applied to differently built lexical networks. The results showed quite dramatic differences, which is not really surprising. After all, different networks are built for different purposes; hence each one of them is more or less suited for a given task. What was more surprising though is the fact that the relational information given by the users did not allow us to find the elusive word in WordNet better than without it.

1   The problem : word access in language production

Communication is largely based on words which encode various sorts of information, *conceptual* (lexical semantics, encyclopedic knowledge), *linguistic* (word forms, part of speech), ... If ever we lack any of this information we may reach for a dictionary, a thesaurus or an encyclopedia in the hope to find what we are looking for. Information access works generally quite well for readers, but much less for authors. Obviously, readers and writers have different needs, and while both provide words as input, they clearly pursue different goals. Readers start from word forms in the hope to get meanings, while authors go the opposite direction: starting from meanings (or meaning fragments), broader topical categories (thesaurus) or specific target-related words (associations, co-occurrences) they hope to find the elusive word (target). We will be concerned here with this latter kind of search.

There are two major access modes, one being automatic, and the other deliberate. The former relies solely on our brain (on-line processing when speaking or writing) whereas the latter uses an additional, external resource (paper or electronic dictionary). In general we resort to this second strategy only if spontaneous access fails. Alas, most dictionaries are not very well suited for this purpose (see Section 3). Yet, even if we had such a dictionary, we are still faced with the problems of input and size. What information shall the user give to allow the resource to guess the elusive word? Since dictionaries are generally quite large, arises the question of how to reduce the entire set of words (scope of the lexicon) to one, the target. This leads to the next question: how to reduce quickly the initial space to a subspace which is neither too big nor too small, that is, how to ensure that the output contains only a reasonable set of candidates (not too big), yet still potentially relevant information? Inconsiderate filtering

might eliminate promising candidates, in which case the space gets too small. To answer these questions, it is interesting to take a look at the ***Tip-of-the-Tongue*** problem (Brown & McNeill, 1996), henceforth ToT.

2   The Tip-of-the-tongue problem

There are different sorts of impairment hindering wordfinding (aphasia, anomia, ...). One of the best known and most intensively studied ones is the ToT-problem (Brown, 1991). Someone is said to be in this state when he knows what to say, he also knows the corresponding form, but for some reason he simply is not able to access it in time, at the very moment of speaking or writing. To get a better understanding of the problem and process at hand, let us replace the task in its natural context, language production.[1] After all, words are generally used in this situation.

   Language production involves three major tasks (Bock, 1996; Levelt, 1989), most of which apply not only for sentence generation, but also for the production of words. Hence, we start from an image or a concept (Level$_1$), which can be very abstract and be linguistically unspecified. Neither nor have information concerning part of speech, or phonology at this level. Indeed, at this stage we may have something like 'move' or 'reptile' but not their concrete lexical forms, for example: 'walk/limp/run'.

   Hence, the speaker must add sufficient information to be able to decide whether, in the case of reptiles, he wants to refer to an 'alligator, 'caiman', or 'crocodile'. These are lexical concepts, i.e. entries in the mental lexicon, also called a *lemmas* (Level$_2$). Note that at this stage we have only an abstract form containing the meaning, part of speech and some general information concerning the phonological form (number of syllables, intonation, …). Yet, it is only at the next step (Level$_3$) that the brain specifies the phonological form, to yield a *lexeme*, the word's concrete form. This allows us then to compute the required motor program to carry out the necessary steps to produce a written or spoken form. A tip-of-the-tongue state occurs if there is an interruption between Level$_2$ and Level$_3$.[2]

   ToT problems can be seen as a puzzle which can be solved by providing or priming the missing elements. This can be done indirectly (cf. Abrams et al., 2007). James and Burke (2000) designed a protocol to do precisely this. They presented some pictures or definitions asking their subjects to find the corresponding word. Those who failed, but knew the word, i.e. those who were in a ToT-state, were used for the main part of the experiment. This group was then divided in two equal parts. Half of participants were asked to read aloud a list of words that cumulatively contained all of the syllables of the ToT word. Suppose someone failed to retrieve the target *abdicate*, in this case he would be asked to read the following list of ten words, *abstract, indigent, truncate, tradition* and *locate,* each of which contains a syllable of the target. The other half was also given a list of 10 words, but phonologically unrelated. Having done this exercise, participants were asked to try again to retrieve the target. And this time most of the members of the group being exposed to phonologically related words succeeded, while the other group did not.

   Obviously, in a natural situation we can neither wait for the phonological primes to occur, nor can we provide them as James et al. did, as this would require knowledge of the target. Yet if we knew the target then we would give it, since this is what the author is looking for. To conclude, we cannot provide the missing parts or offer form-related cues (for example, phonological cues), what we can do though is to provide semantically related words, associations, i.e. words related to the user input.

3   Related work

Concerning lexical access, several communities are concerned: engineers from the natural language generation community (NLG), psychologists, computational linguists and lexicographers. Space constraints prevent us from referring to all this work. Hence, we will focus here mainly on the work

---

[1] For a broad view from a psycholinguistic or neuroscientist's perspective, see (Levelt, 1989; Rapp and Goldrick, 2006; Goldrick, et al., 2014). The equivalent, but from an engineering point of can be found in (Dale & Reiter, 2000, Krahmer and Theune, 2010). For a recent state of the art paper see (Bateman and Zock, 2016).

[2] Levelt's word production model (Levelt et al., 1999) is actually quite a bit more sophisticated. It requires the following six steps : (1) *conceptual preparation* → lexical concept ; (2) *lexical selection* (abstract word) → lemma; (3) *morphological encoding* → morpheme ; (4) *phonological encoding* (syllabification) → phonological word; (5) *phonetic encoding* → phonetic gestural code ; (6) *articulation* → sound wave. Note that it postulates two knowledge bases: the mental lexicon, vital for lemma retrieval, and the syllabary, important for phonetic encoding.

done in lexicography. Note though, that the problem addressed by the NLG community deals with 'lexical choice', but not with 'lexical access'. Yet, before choosing a word, one must have accessed it.

How words are stored and processed in the human mind has extensively been dealt with by psychologists (Aitchinson, 2003; de Deyne and Storms, 2015; deDeyne et al. 2016). Yet, while there are many papers dealing with the *tip-of-the-tongue phenomenon* (Brown & McNeill, 1996), or the problem of lexical access (Levelt et al. 1999), they do not consider the use of computers for helping people in their task (our goal).

Lexicographers bridge this gap. Unfortunately, until recently most of their tools have been built for the language receiver. Nevertheless, nowadays there are also some *tools* for the language producer. For example, Roget's thesaurus (Roget, 1852) or its modern incarnation built with the help of corpus linguistics (Dornseiff, 2003). There are also the Language Activator (Summers, 1993), the Oxford Learner's Wordfinder Dictionary (Trappes-Lomax, 1997), and various network-based lexical resources: *WordNet,* henceforth WN (Miller,1990), *Framenet* (Fillmore et al. 2003); *MindNet* (Richardson et al., 1998), and *HowNet* (Dong & Dong, 2006;). Finally, there are collocation dictionaries (Benson et al., 2010), and web-based tools like Lexical FreeNet[3] or Onelook (Beeferman, 2003), which, like BabelNet (Navigli & Ponzetto, 2012) combines a dictionary (WN) and an encyclopedia (Wikipedia), though putting the emphasis on onomasiological search, access by meaning. Reverse dictionaries have been built by hand (Bernstein, 1975) and with the help of machines (Dutoit and Nugues, 2002). In both cases, one draws on the words occurring in the definition. Thorat and Choudhari (2016) try to extend this idea by introducing a distance-based approach to compute word similarity. Given a small set of words they compare their approach with Onelook and with dense-vector similarity. While we adopt part of their methodology in our evaluation scheme, we are more reserved with respect to their architecture. Since it requires a fully computed similarity matrix for the entire vocabulary, their work cannot scale up: it is unreasonable to assume that the lexicon is stored in a fully connected similarity matrix, which grows quadratically in the size of the vocabulary. Note that while dense representations are easily compared, proximity search is not. It is computationally simply too expensive.

As one can see, a lot of progress has been made during the last two decades, yet more can be done especially with respect to indexing (organization of the data) and navigation.

4   Navigation, a fundamentally cognitive process

As we will show in this section, navigation in a lexical resource is above all a knowledge-based process. Before being able to use a word, we must have acquired it. It is only then that it has become part of our knowledge. Yet, storage does not guarantee access (Zock & Schwab, 2011). This fact has not received the attention it deserves by lexicographers. Note also that there are several kinds of knowledge: declarative, meta-knowledge (not necessarily linguistic) and knowledge states.

- *Declarative knowledge* is what we acquire when learning words (meaning, form, spelling, usage), and this is the information generally encoded in dictionaries. Obviously, in order to find a word or to find the information associated with it, they must be stored, though this is not enough.

- Next, there is *meta-knowledge,* which also needs to be acquired. Being generally unavailable for in(tro)spection, meta-knowledge reveals itself in various ways. For example, via the information available when we *fail to access* a *word* (Schwartz, 2006), or via the *query* we provide at the moment of launching a search. As word association experiments have shown (Aitchison, 2003) words always evoke something. Since this is true for all words one can conclude that all words are connected in our mind, which implies that all words are accessible from anywhere like in a fully connected graph.[4] All we have to do is to provide some input (source word, available information) and follow then the path linking this input to the output (target). Interestingly, people hardly ever start from words remotely related to the target. Quite to the contrary, they tend to start from a more or less direct neighbor of the target, the distance between the two, exceeding rarely the distance of 2.[5]

---

[3] http://www.lexfn.com

[4] Note that this does not hold for WN, as WN is not a single network, but a set of networks. There are 25 for nouns, and at least one for all the other parts of speech

[5] This is probably one of the reasons why we would feel estranged if someone provided as cue 'computer', while his target 'mocha'. The two are definitely not directly connected, though, there is a path between them, eventhough it is not obvious

Also, dictionary users often know the type of relationship holding between the input (prime) and the target. These two observations clearly support our idea that people have a considerable amount of (meta-) knowledge concerning the organization of words in their mind, i.e. their mental lexicon.

- The idea of relationship has been nicely exploited by WN, which due to this feature keeps the search space, i.e. a set of candidates among which the user has to choose, quite small. The idea of relatedness has led lexicographers already in the past to build thesauri, collocation- and synonym dictionaries. Obviously an input consisting only of a simple word is hard to interpret. Does the user want a more general/specific word, a synonym or antonym? Is the input semantically or phonetically related to the target, or is it part of the target word's definition (dog-animal)? In each case the user is expecting a different word (or set of words) as output. Hence, in order to enable a system to properly interpret the users' goals we need this kind of metalinguistic information (neighbor of the target, i.e. source word + relation to the target) at the input.[6] If ever the user cannot provide it, the system is condemned to make a rough guess, presenting all directly connected words. Obviously, such a list can become quite large. This being so, it makes sense to provide the system this kind of information to produce the right set of words, while keeping the search space small.

- ***Knowledge states***, refer to the knowledge activated at a given point in time, for example, when launching a search. What has been primed? What is available in the user's mind? Not all information stored in our mind is equally available or prominent anytime. The fact that peoples' *knowledge states* vary is important, as it co-determines the way a user proceeds in order to find the information he is looking for. This being so, it is important to be taken into consideration by the system designer. In conclusion, all this knowledge must be taken into account as it allows us to determine the search space, reducing its scope, which otherwise is the entire lexicon.

The example here below illustrates to some extent these facts with regard to wordfinding in an electronic resource. Suppose you are looking for a word conveying the idea of a *large black-and-white herbivorous mammal of China*. Yet, for some reason you fail to retrieve the intended form, *Panda*, even though you know a lot concerning the target. People being in this state, called the ToT-problem, would definitely appreciate if the information they are able to access could be used to help them find the target. Figure 1 illustrates the process of getting from a visual stimulus to its expression in language via a lexical resource. Given an external stimulus (A) our brain activates a set of features (B) that ideally allow us to retrieve the target form. If our brain fails, we use a fallback strategy and give part of the activated information to a lexical resource (C) expecting it to filter its base (D) in the hope to find the target (panda) or a somehow related word (E). As one can see, we consider look-up basically as a two-step process. At step one the user provides some input (current knowledge) to which the system answers with a set of candidates, at step two the user scans this list to make her choice.

A: perceptual input, i.e. target	B: associated features in the brain	C: input to lexical resource	D: lexical resource	E: output of lexical resource
	type : bear *lives_in* : China *features* : black patches *diet* : eats bamboo	bear China	**a**ardvark ... ... *panda* ... ... theorem ... **z**ygote	*panda* polar bear

Figure 1: Lexical access a two-step process mediated by the brain and an external resource (lexicon).

5 A Framework for Dictionary Navigation

In this section we will try to answer briefly the following three questions: *What should a resource look like to allow for the search described in the figure here above*? *How to build and how to use it*?

(The chosen elements are always underlined.): *computer* → (*Java, Perl, Prolog ; mouse, printer ; Mac, PC*) ; (1) Java → (*island, programming language) ;* (2) Java (island) → (coffee; Kawa Igen); (3) coffee → (Cappucino, Mocha, Latte). Note that 'Java' could activate 'Java beans', a notion inherent to Java, the programming language. In this case it would lead the user directly to the class (hypernym) containing the desired target word (mocha).

[6] This has of course consequences with respect to the resource. To be able to satisfy the different user needs (goals, stratgies) we probably need to create different databases: Obviously, to find a target on the basis of sound (rhymes), meanings (meaning-fragments) or related words (co-occurrences), requires networks encoding a different kind of information.

(a) *What should a resource look like to allow for this*? We would need a fully connected graph, or, more precisely, an association thesaurus (AT) containing typed and untyped links. Both kinds of links are necessary for filtering, i.e. to ensure that the search space is neither too big (typed links), nor too small (untyped links). Untyped links are a necessary evil: they are necessary to address the fact that two words evoke each other eventhough we are not able qualify the nature of the link.

(b) *How to use it*? Imagine an author wishing to convey the name of a beverage typically served in coffee shops. Failing to evoke the desired form ('mocha'), he reaches for a lexicon. Since dictionaries are too huge to be scanned from cover (letter A) to cover (Z), we will try to reduce the search space incrementally. Having received some input from the user, say 'coffee', — which is the word coming to his mind while failing to access the target,— the system answers with a set of words among which the user chooses. Iff the input and the target are direct neighbors in the network, and iff the user knows the link between the two (source + target), then the search space is generally quite small. In the opposite case, that is, if the user cannot specify the link, then the system is condemned to make an exhaustive search, retrieving all direct neighbors of the input. However, the system could cluster the words by affinity and give names to these categories, so that the user, rather than navigating in a huge flat list navigates in a *categorial tree*, which avoids scanning long lists.

(c) *How to build it*? While there are quite a few resources, in particular, association thesauri, they are too small to allow us to solve the ToT-problem. Projected resource would still have to be built, and while one could imagine the use of combined resources, like Babelnet (Navigli and Ponzetto, 2012), or the combination of WN with other resources like topic maps (Agirre et al. 2001), Roget's Thesaurus (Mandala, 1999) or ConceptNet (Liu and Sing, 2004), it is not easy to tell which combination is best, all the more as besides encyclopedic knowledge, we also need episodic knowledge (Tulving, 1983).

One straightforward solution might be co-occurrences (Wettler & Rapp, 1993; Lemaire & Denhière, 2004; Schulte im Walde & Melinger, 2008). While co-occurring words contain many appropriate clue – target pairs, they also contain many unrelated terms that hamper access – even after application of appropriate significance measures. More severely, there are no structural elements that generalize across queries.

Another solution could be *lexical functions* (Mel'čuk, 1996) or semagrams (Moerdijk, 2008) which are reminiscent of the lexical-semantic networks produced by Fontenelle (1997) on the basis of the Collins-Robert dictionary enriched with Melcuk's lexical functions. Semagrams represent the knowledge associated with a word in terms of attribute-values. Each semantic class has its type template and corresponding slots. For instance, the type template for *animals* contains the slots 'parts, behavior, color, sound, size, place, appearance, function', etc., whereas the one for *beverages* has slots for 'ingredient, preparation, taste, color, transparency, use, smell, source, function, 'composition', etc. While it is unlikely that we can infer or mine semagrams automatically, chances are that we can populate them mechanically, which would then be seen as an alternative route of building an association thesaurus, but in a fairly controlled way.

6 Experimental Setup

In this section, we describe the experimental set-up to answer the following research questions: (a) *When being in the ToT-state what cues do people provide to help the system find the target?* (b) *How good are existing lexical resources for retrieving the targets by using these cues?* (c) *How big is the added value of knowing the relationship between the cue (source word) and the target?* Put differently, does it enhance retrieval precision and speed?

6.1 Lexical Graphs as Dictionaries

For our experiments we used three different lexical networks: WN, distributional semantic models using word similarity and word co-occurrence. They were chosen deliberately to cover different structural aspects, different amounts of effort to construct them manually, and different degrees of language-dependence. Note, that we could have chosen other resources, for example, the Edinburgh Association Thesaurus,[7] but the E.A.T lacks typed relations and it is quite old (Kiss et al. 1973),[8] covering only a subset of the words used in our experiment.

---

[7] Available at: http://www.eat.rl.ac.uk

- *WordNet:* WN 3.0 (Fellbaum, 1998) is a high-coverage, manually built lexical-semantic network of English. Words are organized in terms of synsets, i.e. sets of synonyms, which are linked in various ways depending on the part of speech. We used a subset of these links (synset, hyponymy, derivation, etc.) and domain categories in the hope to be able to retrieve the target.

- *Word Similarity:* We used the JoBimText distributional semantic model, its similarity score being based on common dependency parse contexts, which requires a language-specific parser. The JoBimText distributional thesaurus[9] (Biemann and Riedl, 2013) contains in ranked order the 200 most similar terms of a newswire corpus of 100 million sentences in English. We expect this resource to be suitable for most associative queries, that is to help us find words occurring in contexts like "X is somehow like a Y or a Z" (e.g. "a *panda* is somehow like a *koala* or a *grizzly*"). This example illustrates 'co-hyponymy', a relation not directly encoded in WordNet. Similarities (for example, panda/koala vs. panda/dog) are ranked by context overlap.

- *Word Co-occurrence*: We compute statistically significant sentence-based word co-occurrences using the same corpus as here above, and following the methodology of (Quasthoff et al., 2006)[10]. We expect this resource to be suited for free associations, i.e. cue words whose link to the target cannot be specified. This resource has by far the highest rate of relations across different word classes, as they may occur in patterns like "With Xs, especially with Y ones, you can Z and W" (e.g. "with mochas, especially with iced ones, you can chill and have cookies"). Co-occurrences are ranked by the log-likelihood significance measure (Dunning, 1993).

6.2 Network Access

Given the structural differences of our resources, our networks are accessed with different query strategies. The general setup is to query the resource via a cue and to insert then the retrieved terms into a ranking. As long as the system has not found all the desired words, it will keep going by querying with words according to their rank, inserting previously un-retrieved terms below the ranking.

- *WordNet*: Having noticed that people tend to use *hypernyms* (flower) as cues to find the *hyponym* (rose, the target), we defined a heuristic supporting queries using this relation. We start by querying for 'synonyms' of the cue, putting results first in the ranking. Next, we proceed along the sense numbers, senses being ordered by frequency in WN, which ensures that we start with the most common senses. Third, we add (in this order) direct 'hyponyms', 'meronyms' and 'domain members'. This order seems to be justified by the fact that most people tend to go from general to specific, starting by a more general term when launching a search. Finally, we add other relations like 'similar', 'antonyms', 'hypernyms', 'holonyms, 'domains', etc. For example, for the cue "pronouncement", the target "affirmation" is found by first checking the cue's 'synonyms' ("dictum", "say-so"), before checking the direct *hyponym* and *hypernym* (directive, declaration). Next we navigate through directly related words of "dictum", synonym of "pronouncement", to find then the target as a direct *hypernym* of "say-so" in its first sense, resulting in rank 12.

- *Word Similarity*: We retrieve the most similar terms per query, ranked by their similarity. Note that due to structure limitations of the resource only 200 similar words can be retrieved per query.

- *Word Co-occurrence*: Having filtered out the 200 most frequent stopwords, we retrieve terms co-occurring at least twice with a minimum log-likelihood score of 6.63.

Each cue returns a ranking of the full vocabulary. Working with three cues per target (see Section 6.3), we explore two different combinations of target ranks (minimum rank and merged rank) from querying with the three cues. Regarding *minimum rank*, the rationale is that for each cue, a retrieval process

---

[8] For example, if you provide 'terrorism' as key, you will get the following list of ranked words as answer : Guerilla, Gun, Soldier, War, Guerrilla, Anarchist, Evil, Fear, Fighting, Rebel, Tyrant, Vandal, Vietnam, Abroad, Activities, Activity, Arab, Arson, Bandit, Blood, Bomb, Che, Che Guevara, Congo, Czech, Fight, Fighter, Gangster, Gorilla, Greek, Guerillas, Guns, Hooligan, Kill, Killer, Madness, Man, Mao, Maoist, Mexico, Night, Police, Regime, Revolution, Revolutionary, Rioter, Russian, Shoot, Terror, Tourist, Tree, Trotsky, Vietcong, Vietnamese, Wog. As one can see 'associations' change over time. The words we would associate nowadays with 'terrorism' are not the same as the ones people had associated in the seventees, the moment of history where this resource was built.

[9] Available at www.jobimtext.org

[10] Available at http://corpora.informatik.uni-leipzig.de

is started in parallel, terminating when the ToT target is encountered for the first time. Actually, only the rank of the 'best' cue is used. For *merged rank*, the rationale is as follows: we use all cues and merge the three rankings by a) adding the ranks per word and sorting by sum or b) multiplying the ranks and sorting by product. For more details, see Section 6.4.

6.3   Dataset

Since it is not trivial to put people in the ToT state, we have reformulated the problem in the following way: we ask people to describe a given target to other people who may not know the word (e.g. language learners), by providing three cues. Crowdworkers were asked to provide single-word cues rather than descriptions or definitions. Note that the idea was not the creation of a resource, but rather the creation of a set of data to see how well they would behave with respect to our three resources (section, 6.1). Also, in order to get a clearer picture concerning our third question, i.e. the added value of the relation between cue and target, we asked subjects to also specify the relationship between the target and each one of the given three cues. Relations were defined indirectly, i.e. via examples. They comprise synonyms, hypernyms/hyponyms, meronyms/holonyms, typical properties, typical roles (verb-subject, verb-object) and free associations.

Data acquisition was done via the Crowdflower crowdsourcing platform.[11] In order to check whether crowdworkers had given the right answer and understood the target, we presented the latter together with three definitions. For our experiment we used only trials that the crowdworkers had fully understood, that is, for which they had picked the correct definition. After data collection, we excluded data from crowdworkers that deliberately had ignored our instructions. For the targets and definitions we used the 208 common nouns listed in (Abrams et al., 2007; Harley and Bown, 1998), who examined the ToT state from a psychological angle. Full data, instructions and judgments are available online.[12]

Data collection yielded a total of 1186 cue triplets, provided by 65 participants, who worked on 3 to 132 targets. After manual correction of typos and lemmatization, cue triplets were filtered by eliminating words outside of the vocabulary of the respective resource used in the experiments. Inspection of the data revealed that crowdworkers generally chose the cues quite well, but many of them had a hard time to assign the appropriate relation, which is not all that surprising, as this requires quite a bit of metalinguistic knowledge. It is also possible that some participants had chosen the relation without taking the needed care since we did not perform any quality checks during the task. We probably need a different kind of experiment to validate this or measure the extent to which linguistically innocent users can accurately classify semantic relations.

Table 1 below shows the distribution of relations expressed in the first 200 cue triplets (target range 'a-c', i.e. abacus – calisthetics, in alphabetical order) containing also some manually assigned relations. The results show the importance of taxonomic relations, a fact well exploited by WN. Representing nearly 46% of the relations, they confirm the intuition that paradigmatic associations are an important means to access the desired word. However, the next largest class are *syntagmatic*, i.e. untyped, associations (37%). Note that about 17% of the cues come from a different word class than the targets.

Relation	associated	hyponym	synonym	quality	object	meronym	holonym	subject	hypernym
Ex.: cue - target	tea - afternoon	story - anecdote	horoscopy - astrology	white - albatross	share - anecdote	letters - anagram	day - afternoon	cheer - audience	zombie - cadaver
Typ. POS	N	N	N	A	V	N	N	V	N
%	36.8%	23.5%	13.3%	8.2%	5.2%	4.3%	4.2%	3.8%	0.6%

Table 1: Distribution of relations between target and cue, as well as typical part of speech (POS) for the cue (N: *Noun*, V: *Verb*, A: *Adjective*), manually assigned by the authors.

6.4 Evaluation Methodology

Our methodology is very similar to the one of Thorat and Choudhari (2016): we query the lexical network with cues and retrieve then a ranked list of potential ToT targets. With more appropriate cues and better lexical resources, our targets will probably get a boost, appearing higher in the list.

Our vocabulary of WN comprises 139,784 terms, including multiwords, which can be mutually reached through the query procedure described above was used in the first experiment. The intersection of the vocabulary of the three networks consists of 34,365 terms, all of them being single words, just as the ones used in the second experiment. Here below are the criteria used in our evaluation:

- *Minimum rank per cue (MinRank)*: if all cues were processed strictly in parallel, when would the target appear for the first time?

- *Target rank in sum of ranks (+Rank)*: if the retrieval time depends on the average rank per cue, we sum the ranks of the three cues and sort the list of terms in ascending order, reporting the position of the target. Note that this score is strongly influenced by negative outlier cues.

- *Target rank in multiplication of ranks (*Rank)*: To model a multiplicative instead of an additive combination, we multiply the target ranks per cue, sort the list of terms by this score in ascending order, and report then the position of the target. This score is less sensitive to negative outliers.

- *Average Precision@100 (P@100)* measures the fraction of trials containing the target among the first 100 hits, for each of the above. While 100 is an arbitrary number, it seems a reasonable wordlist size to allow for the quick retrieval of a target.

Note that the *minimum rank* is not necessarily lower than the other two scores. It is possible, and it even happens in our data, that a target gets a low rank because all three cues rank it consistently low, while the targets preferred by single cues are ranked much less favorably than others. For example, the target "agnostic" was retrieved from WN (untyped) by its three cues "believer, god, atheist" with ranks 170, 890, respectively 25. Minimum Rank is thus 25, but ranking via sum of ranks lists the target at position 14, while the multiplicative combination results in rank 15.

In the next section, we will qualitatively assess the differences in rankings from our different semantic networks.

7 Results and Discussion

We ran two experiments. In the first we tried to find out whether the knowledge and usage of WN relations produces some added value in terms of retrieval. The goal of the second experiment was to compare the retrieval performance of our three dictionary resources.

7.1 Retrieval along Semantic Relations

To answer the question whether the usage of relations improves word access, i.e. retrieval, we used WN, as it is highly structured and our relations can be directly mapped to it. For incorporating relations, we adapted the following query procedure (cf. Section 6.2): we first query for the target relation and then for all the others. For example, for the target "abacus" and the clue "bead" of type *meronym*, we would first retrieve the *holonyms* of "bead", then all other relations in the order given in Section 6.2, for initial and subsequent queries. If the supplied relation between the cue and the target is directly given in WN, retrieval is quick. Since the WN hierarchy is quite fine-grained, and since a hyponym relation might be contained over several transitive steps, we keep this order throughout the entire query process.

strategy \ score	MinRank	P@100	+Rank	% top100	*Rank	P@100
WordNet untyped	12352.7	40.5%	**22403.2**	7.5%	17993.5	21.5%
WordNet relations	**11733.2**	**42.0%**	22722.7	**9.5%**	**17786.0**	**22.5%**
Random Baseline (STDEV)	35480.7 (514.8)	0.2%	70264.7 (636.0)	0.1%	70438.5 (777.1)	0.1%

Table 2: Scores for target retrieval in WordNet by using or ignoring relational information for 200 cue triples on a vocabulary of 139,784 terms

Both settings perform much better than the random baseline, which returns the vocabulary in random order irrespective of the dictionary's structure. The random baseline was obtained by running

simulations over the same size of the dataset; we also provide the standard deviation on 10 runs in parenthesis where applicable. Since more than 40% of the targets are among the first 100 retrieved words in the MinRank setting, we conclude that WN is indeed suitable. A manual analysis statistically confirmed our intuition: WN is very good for retrieving targets based on taxonomically related cues (e.g. calculator – abacus), while it does not perform well at all for syntagmatically related words or for noun-noun cues (e.g. beads – abacus, gluten – allergy).

Regarding the added value of relations for retrieval, we conclude that typed relations only help to a small extent, if at all. Our data show fluctuations in the range of a relative -2% to +5% between the settings. Note that this may be a side effect of the sample size, which is quite small. Interestingly, the differences decreased when repeating the experiment with the smaller vocabulary from Experiment 2. Clearly, more work is needed here.

7.2 Comparison of the three Resources

In order to assess differences between our dictionary resources, we consider the 964 cue triplets per target matching, the common vocabulary of our three resources (see Table 3 below).

dictionary \ score	MinRank	P@100	+Rank	P@100	*Rank	P@100
Word Similarity	**523.6**	**61.0%**	**1945.9**	**40.6%**	**1040.2**	**55.7%**
Word Co-occurrence	1748.0	44.2%	4205.6	27.2%	3226.9	33.6%
WordNet	2615.4	51.2%	6132.9	13.0%	4247.2	30.3%
Random Baseline (STDEV)	8543.0 (189.7)	0.9%	17156.6 (260.7)	0.2%	17113.8 (252.0)	0.3%

Table 3: Scores for target retrieval in our resources for 964 cue triples
based on a common vocabulary of 34,365 words.

All dictionaries allow for much better retrieval than the random baseline. The results provide a clear picture: the *word similarity* resource achieves the lowest average ranks on all scores. In 61% of the cases, the target is among the top 100 retrieved words if we consider only the most effective cue (MinRank). Note that more than half of the targets are found in the top 100 for the multiplicative combination (*Rank). This is surprising, as the relations between the cues and the target are quite diverse (see Section 6.3), and Word Similarity mostly contains direct and indirect taxonomic relations, such as co-hyponyms. The second-best resource in this evaluation is the word co-occurrence network, which outperforms WN on all metrics except the P@100 of MinRank scores.

We also analyzed the differences qualitatively and looked at cue-target-pairs where the three networks perform very differently. As our findings show, different networks have different potentials with respect to the retrieval of ToT targets based on a given cue:

- *WordNet* good, *Co-occurrence* poor: Synonyms or near-synonyms, like javelin – spear, cadaver – corpse. These do not co-occur in sentences, also cf. (Biemann et al., 2012).

- *WordNet* poor, *Co-occurrence* good: associations, like hospital–doctor or hospital–sick. They are not encoded in WordNet, its associative relations are very spotty. Note that placing them first in the order of relations did not increase performance.

- *WordNet* good, *Similarity* poor: meronyms/holonyms, such as door–knob, road–asphalt. These are not similar at all from a distributional point of view.

- *WordNet* poor, *Similarity* good: relations that should be in WN, but for some reason are missing, e.g. torpedo–missile, calligraphy–art, gazebo–pavilion.

- *Co-occurrence* good, *Similarity* poor: associations, part-of and cross-POS-relations, such as orthodontist–braces, hospital–ER and growth–economic. Though being related, these words are not similar.

- *Co-occurrence* poor, *Similarity* good: (near) synonyms, such as mercenary–warrior, lampoon–caricature, orthodontist–dentist. Again, they rarely co-occur in the same sentence.

8 Final Comments and Conclusion

In this paper, we have examined the use of lexical semantic networks to overcome the ToT problem. After an analysis of the causes leading to this state, we have evaluated and analyzed three lexical

networks meant to overcome the ToT problem: WordNet, a word similarity network and a word co-occurrence network. Our setup was to query the network with a cue and check whether this would allow us to retrieve the target. To see its relative efficiency, we measured the rank of the ToT target over the retrieved vocabulary.

A ToT state can be induced by describing a given target to another person by providing some cues and ask him then to name it. Something similar can be achieved via crowdsourcing. We assumed that the cues retrieved via this technique, are similar to the ones humans typically use for the target retrieval. In order to determine the added value of a cue, we asked subjects to specify also the relationship between the target and each one of the given three cues. It turned out that traditional X-'onym' relations (hyponym, hypernym, ...) represent about half of the relations, while the remainder are mainly associated terms, i.e. untyped relations.

While we could not successfully exploit relational information to enhance retrieval, we could show the relative efficiency of different lexical semantic networks with respect to word access. As expected, *WordNet* is very good for retrieving targets on the basis of synonyms or taxonomically related cues. *Word co-occurrence* excels in associations, qualities and typical actions. Yet, the best network in our experiment was the one based on *word similarity*, as, apart for meronym/holonym relations, it combines the advantages of the other two. It covers basically the same aspects as WN, but it is more complete, containing syntagmatically associated terms like the co-occurrence network.

The fact that WN does not perform well for syntagmatically related words suggests the usage of another resource like Mel'čuk's *Explanatory Combinatory Dictionaries* (ECD) (Mel'čuk, 2006). ECDs look like good candidates, possibly better suited for our task than WN. Being part of a language production theory, called 'Meaning-Text Model' (Mel'čuk, 2012), ECDs capture a larger range of lexical relations (50+ lexical functions) than WN. Alas, the problem we have with this option are coverage and availability. Though being extremely fine-grained the ECD covers so far only a subset of the words normally found in a lexicon. Also, the ECD is not available in digital form.

Other potentially interesting alternatives would be association networks. Unfortunately, these resources are either not free (Gavagai),[13] too old (Kiss, et al. 1973), not rich enough in terms of coverage (de Deyne, et al. 2016; Nelson, et al. 2004), or not in the needed language, English (Lafourcade, 2007, 2015). Probably the largest, and arguably the best association thesaurus at this moment is JeuxDeMots, a crowd-sourced resource created via a game, hence its name 'wordgames'.[14] Yet, as mentioned already, this resource is not available in English, which is probably the reason why it is so little known 'abroad'.

One last word concerning 'relations'. Since we do believe in the virtues of relational information, —they are a critical component of the input— we plan to re-visit the problem of navigation in lexical graphs, but on the basis of cues enriched with relational information. Relations provide a context for the input. Revealing the users' goal, they tell the information provider (human or system) what to do with the input: provide a synonym, hypernym, etc. Obviously, a user expects quite different outputs for the following inputs : ['similar_to' 'knife'], ['more general' than 'knife'], or ['part_of' 'knife']. Since our ultimate goal is the creation of a resource helping people to overcome the ToT problem, we plan to combine different types of corpora, to build then a hybrid semantic network, that is, an association thesaurus containing typed and untyped relations. The first to keep the search space small, the second to make it large enough to include potentially relevant words, possibly even our target.

References

Abrams, L., Trunk, D. L., and Margolin, S. J. (2007). Resolving tip-of-the-tongue states in young and older adults: The role of phonology. *In L. O. Randal (Ed.), Aging and the Elderly: Psychology, Sociology, and Health* (pp. 1-41). Hauppauge, NY: Nova Science Publishers, Inc.

Agirre, E., Ansa, O., Hovy, E., and Martinez, D. (2001). Enriching WordNet concepts with topic signatures. *In: SIGLEX workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations*, Pittsburgh, PA, USA.

---

[13] https://explorer.gavagai.se and https://lexicon.gavagai.se
[14] http://www.jeuxdemots.org/jdm-accueil.php and http://www.jeuxdemots.org/AKI.php

Aitchison, J. (2003). *Words in the Mind: an Introduction to the Mental Lexicon*. Oxford, Blackwell. (original version, 1987)

Bateman J. and M. Zock. (2016) Natural Language Generation. In: *R. Mitkov (Ed.) Handbook of Computational Linguistics (2nd edition)*, Oxford University Press. Forthcoming.

Beeferman, D. (2003). Onelook reverse dictionary. http://onelook.com/reverse-dictionary.shtml.

Bernstein, T. (1975). *Bernstein's Reverse dictionary*. Crown, New York.

Biemann, C., Riedl, M. (2013). Text: Now in 2D! A Framework for Lexical Expansion with Contextual Similarity. *Journal of Language Modelling* 1(1):55-95.

Biemann, C., Roos, S., and Weihe, K. (2012). Quantifying Semantics Using Complex Network Analysis. In: *Proceedings of COLING-12*, Mumbai, India, pp. 263-278.

Bock, J.K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review*, 3:395-421.

Brown, A. S. (1991). A review of the Tip-of-the-Tongue Experience. *Psychological Bulletin*, 109:204 – 223.

Brown, R and Mc Neill, D. (1966). The tip of the tongue phenomenon. *Journal of Verbal Learning and Verbal Behaviour*, 5:325-337.

de Deyne, S., and Storms, G. (2015). Word associations. In J. R. Taylor (Ed.), *The Oxford Handbook of the Word*. Oxford University Press, Oxford, UK.

de Deyne, S., Verheyen, S. and Storms, G. (2016). Structure and organization of the mental lexicon: A network approach derived from syntactic dependency relations and word associations. In: *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks* (pp. 47-79). Springer Berlin Heidelberg.

Dong, Z. and Q. Dong. (2006). *HOWNET and the computation of meaning*. World Scientific, London.

Dornseiff, F. (2003). *Der deutsche Wortschatz nach Sachgruppen*. Berlin & New York: W. de Gruyter.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74.

Dutoit, D. and P. Nugues (2002): A lexical network and an algorithm to find words from definitions. In: *van Harmelen, F. (ed.): Proceedings of the 15th European Conference on Artificial Intelligence*, pp.450-454 Lyon, France.

Fellbaum, C. (Ed.) (1998). *WordNet: An electronic lexical database and some of its applications*. MIT Press.

Fillmore, C., Johnson, C., and Petruck, M. (2003). Background to FrameNet. *International Journal of Lexicography* 16:235–250.

Fontenelle, T. (1997). Using a bilingual dictionary to create semantic networks, International Journal of Lexicography, Vol.10, n°4, Oxford University Press, pp.275-303

Goldrick, M. A., Ferreira, V., and Miozzo, M. (2014). *The Oxford handbook of language production*. Oxford University Press.

Harley, T. A., and Bown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology*, 89:151–174.

James, L., and Burke, D. (2000). Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26:1378-1391.

Kiss, G.R., Armstrong, C., Milroy, R., and Piper, J. (1973). An associative thesaurus of English and its computer analysis. In: *A. Aitken, R. Beiley and N. Hamilton-Smith (eds.): The Computer and Literary Studies*. Edinburgh: University Press. pp. 153-16

Krahmer, E. and Theune, M. (Eds.) (2010). *Empirical Methods in Natural Language Generation-Data-oriented Methods and Empirical Evaluation*. Series: Lecture Notes in Computer Science, Vol. 5790, Springer Berlin Heidelberg.

Lafourcade, M., and Joubert, A. (2015). TOTAKI: A help for lexical access on the TOT Problem. In Gala, N., Rapp, R. et Bel-Enguix, G. (Eds). *Language Production, Cognition, and the Lexicon*. Festschrift in honor of Michael Zock. Series Text, Speech and Language Technology XI. Dordrecht, Springer, pp. 95-112

Lafourcade, M. (2007). Making people play for Lexical Acquisition with the JeuxDeMots prototype. In: *Proceedings of the 7th International Symposium on Natural Language Processing*, Pattaya, Chonburi, Thailand.

Lemaire, B. and Denhière, G. (2004). Incremental construction of an associative network from a corpus. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society (CogSci'2004)*, 825-830

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

Levelt, W. (1989). Spe*aking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42:1-22.

Liu, H. and Singh, P. (2004). ConceptNet: A practical commonsense reasoning tool-kit. *BT Technology Journal* 22(4):211–226

Mandala, R., Tokunaga, T. and Tanaka, H. (1999). Complementing WordNet with Roget's and Corpus-based Thesauri for Information Retrieval. In: *Proceedings of EACL 99*, pp. 94-101, Bergen, Norway

Mel'čuk, I. (1996). Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. In L. Wanner (ed), Lexical Functions in Lexicography and Natural Language Processing. Language Companion Series 31, Amsterdam/Philadelphia: John Benjamins, 37–102.

Mel'čuk, I. (2006). Explanatory Combinatorial Dictionary. In G. Sica (ed), Open Problems in Linguistics and Lexicography. Monza: Polimetrica, 225–355.

Mel'čuk, I. (2012). Semantics: From meaning to text. Volume 1, Studies in Language Companion Series 129, Amsterdam/Philadelphia: John Benjamins.

Miller, G.A. (ed.) (1990). WordNet: An On-Line Lexical Database. *International Journal of Lexicography*, 3(4):235-244.

Moerdijk F. (2008). Frames and semagrams; meaning description in the general Dutch dictionary. In: *Proceedings of the Thirteenth Euralex International Congress, EURALEX*, pp. 561-569, Barcelona, Spain.

Navigli, R. and Ponzetto, S. (2012), BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. *Artificial Intelligence* 193:217-250.

Nelson, D., McEvoy, C., and Schreiber, T. (2004). The university of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods* 36(3):402–407.

Quasthoff, U., Richter, M. and Biemann, C. (2006). Corpus Portal for Search in Monolingual Corpora. *In: Proceedings of LREC-06*, pp. 1799-1802, Genoa, Italy.

Rapp, B. and Goldrick, M. (2006). Speaking words: Contributions of cognitive neuropsychological research. *Cognitive Neuropsychology*, 23 (1):39-73.

Reiter, E. and Dale, R. (2000). *Building natural language generation systems* (Vol. 33). Cambridge: Cambridge University Press.

Richardson S, W., B. Dolan and L. Vanderwende. (1998). MindNet: Acquiring and Structuring Semantic Information from Text. In: *Proceedings of ACL-COLING'98*, pp. 1098-1102, Montreal, Canada.

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

Schwartz, B. L. (2002). *Tip-of-the-tongue states: Phenomenology, mechanism, and lexical retrieval*. Mahwah. New Jersey: Lawrence Erlbaum Associates.

Schwartz, B. L. (2006). Tip-of-the-tongue states as metacognition. *Metacognition and Learning*, 1(2), 149-158.

Schulte im Walde. S., and Melinger, A. (2008). An in-depth look into the co-occurrence distribution of semantic associates. *Rivista di Linguistica, Special Issue on From Context to Meaning: Distributional Models of the Lexicon in Linguistics and Cognitive Science*, 20(1):89-128.

Summers, D. (1993). *Language Activator: the world's first production dictionary*. Longman, London.

Thorat, S., and Choudhari, V. (2016). Implementing a Reverse Dictionary, based on word definitions, using a Node-Graph Architecture. In: *Proceedings of COLING 2016*, Osaka, Japan.

Trappes-Lomax, H. (1997). *Oxford Learner's Wordfinder Dictionary*. Oxford: Oxford University Press.

Tulving E. (1983). Elem*ents of Episodic Memory*. Oxford: Clarendon.

Wettler, M., and Rapp, R. (1993). Computation of word associations based on the co-occurrences of words in large corpora. In: *Proceedings of the 1st Workshop on Very Large Corpora*, pp. 84-93, Beijing, China.

Zock, M. and D. Schwab, (2011). Storage does not guarantee access. The problem of organizing and accessing words in a speaker's lexicon. *Journal of Cognitive Science* 12(3):233-258.