



**HAL**  
open science

# Eureka! A Simple Solution to the Complex ‘Tip-of-the-Tongue’-Problem

Michael Zock

► **To cite this version:**

Michael Zock. Eureka! A Simple Solution to the Complex ‘Tip-of-the-Tongue’-Problem. Bastardas-Boada, A.; Massip Bonet, A; Bel-Enguix, G. Complexity Applications in Language and Communication, pp.251-272, 2019, 10.1007/978-3-030-04598-2\_14 . hal-02079168

**HAL Id: hal-02079168**

**<https://hal.science/hal-02079168v1>**

Submitted on 14 Mar 2021

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

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

Copyright

# Chapter 14

## Eureka! A Simple Solution the Complex ‘Tip-of-the-Tongue’- Problem



Michael Zock

*To search for a word in a dictionary without a proper index is  
like looking for an address in a city without a decent map.*

1 **Abstract** Dictionaries are repositories of knowledge concerning words. While  
2 readers are mostly concerned with *meanings*, writers are generally more concerned  
3 with *word forms* expressing meanings (lemma). I will focus here on this latter task:  
4 building a tool to help authors to find the word they are looking for, word they may  
5 know but whose form is eluding them. Put differently, my goal is to build a resource  
6 helping authors to overcome the Tip-of-the-Tongue problem (ToT). Obviously, in  
7 order to access a word, it must be stored somewhere (brain, resource). Yet this is  
8 far from sufficient. Access may depend on many other factors than storage of word  
9 forms: organization of the dictionary (index), the user’s cognitive state, i.e. available  
10 knowledge at the onset of search, the distance between the source- and the target-  
11 word (direct neighbor or not) , the knowledge of the relationship between the two,  
12 etc. I will try to provide evidence for the claim that (word) storage does not guar-  
13 antee access. To this end I will compare a well-known lexical resource, WordNet  
14 (WN), to an equivalent one, but bootstrapped from Wikipedia (WiPe). While both  
15 versions contain basically the same set of words, the latter contains many more  
16 (syntagmatic) links than WN. This is probably the reason why WiPe outperforms  
17 WN. In the last two sections I will explain under what conditions WN is suitable  
18 for word access, and what it might take to go beyond the limitations of this famous  
19 resource.

---

M. Zock (✉)  
Aix-Marseille Université, CNRS, LIF (UMR 7279), 163, Avenue de Luminy,  
13288 Marseille, France  
e-mail: [michael.zock@lis-lab.fr](mailto:michael.zock@lis-lab.fr)

© Springer Nature Switzerland AG 2019  
À. Massip-Bonet et al. (eds.), *Complexity Applications in Language  
and Communication Sciences*, [https://doi.org/10.1007/978-3-030-04598-2\\_14](https://doi.org/10.1007/978-3-030-04598-2_14)

1

## 20 14.1 Introduction

21 Speaking a language can be a daunting task. Planning *what to say* (message) and  
 22 *how to say it* (linguistic form) have to be carried out on the fly, that is, quasy simul-  
 23 taneously, and while finally *saying* it (articulation) one may have to plan already  
 24 the next stretch of discourse. Actually, speaking is quite a bit more complex than  
 25 that, requiring the solution of at least half a dozen of problems: determine content,  
 26 find suitable words and sentence frames, add function words, perform morphological  
 27 operations (agreement) and articulate.

28 To get a better idea of the complexity of the task at hand let's focus only on one of  
 29 them, lexical access. Bear in mind though that speech is fast (2–5 words per second,  
 30 Levelt 1993, 2001), and that words must be found in a resource containing, say,  
 31 100,000 words (Miller 1991). This raises a number of interesting questions:

- 32 • how do people manage to access words so quickly in such a huge lexicon, suc-  
 33 ceeding most of the time while making very few mistakes?
- 34 • can the insights gained by studying the *mental lexicon* (storage, organization,  
 35 process) be transposed to an external resource? Put differently, can we use this  
 36 knowledge to enhance paper- or electronic dictionaries (off-line processing)?
- 37 • If ever the answer is negative. Are there ways to achieve something equivalent, at  
 38 least in terms of precision, that is, can we build or enhance an existing resource  
 39 (electronic dictionary) in such a way as to allow its users to find quickly the desired  
 40 word?

41 Having addressed the first two questions elsewhere (Zock et al. 2010), I will focus here  
 42 mainly on the last problem, building a resource meant to help users to overcome the  
 43 *tip-of-the-tongue problem*.<sup>1</sup> Hence, functionally speaking I try to achieve something  
 44 equivalent to the human brain, though in slow motion: help people to find the word  
 45 they are looking for. Before showing the way how this can be done, let me say a few  
 46 words concerning the word-access problem.

47 When speaking or writing we encounter basically either of the following two  
 48 situations: one where everything works automatically (Segalowitz 2000), somehow  
 49 like magic, words popping up one after another as in a fountain spring, leading to a  
 50 discourse where everything flows like in a quiet river (Levelt et al. 1999; Rapp and  
 51 Goldrick 2006). The other situation is much less peaceful: discourse being hampered  
 52 by hesitations, the author being blocked somewhere along the road, forcing her to  
 53 look deliberately and often painstakingly for a specific, possibly known word (Zock  
 54 et al. 2010; Abrams et al. 2007; Schwartz 2002; Brown 1991). I will be concerned here

---

<sup>1</sup>The ToT problem is characterized by the fact that the author has only partial access to the word form s/he is looking for. The typically lacking parts are phonological (Aitchison 2003). The ToT problem is a bit like an incompleated puzzle, containing everything apart from some minor small parts (typically, syllables, phonemes). Alas, not knowing what the complete picture (target, puzzle) looks like, we cannot determine the lacking part(s). Indeed, we cannot assume to know the target, and claim at the same time to look for it or any of its elements. Actually, if we knew the target (word) there wouldn't be a search problem to begin with, we would simply spell out the form.

55 with this latter situation. More specifically, I am concerned here with authors using  
 56 an electronic dictionary to look for a word. While there are many kind of dictionaries,  
 57 most of them are not very useful for the language producer. The great majority of  
 58 them are *semasiological*, that is, words are organized alphabetically. Alas, this kind  
 59 of organisation does not fit well the language producer whose *starting points* (input)  
 60 are generally meanings<sup>2</sup> or cue-words (primes) and only the *end point* (outputs) the  
 61 corresponding target word.

62 To be fair though, one must admit that great efforts have been made to improve  
 63 the situation both with respect to lexical resources and electronic dictionaries. Since  
 64 the invention of the thesaurus (Roget 1852) quite a few *onomasiological dictionaries*  
 65 have been built (van Sterkenburg 2003; Casares 1942), even if nowadays they are not  
 66 built by hand anymore (Dornseiff et al. 2004; Rundell and Fox 2002). Today we build  
 67 lexical resources via corpora (Hanks 2012; Kilgarriff and Kossem 2012), crowd-  
 68 sourcing (Benjamin 2014), dictionary writing systems (Abel 2012) and corpus tools  
 69 like Sketch Engine (Kilgarriff et al. 2004), MonoConc or WordSmith. Yet thesauri  
 70 are not the only kind of onomasiological resources. There are *analogical dictionaries*  
 71 (Boissière 1862; Robert et al. 1993), *collocation dictionaries* (Benson et al. 2010),  
 72 *reverse dictionaries* (Bernstein 1975; Kahn 1989; Edmonds 1999), *rhyme dictio-*  
 73 *naries* (Fergusson and Fergusson 1985 ; Webster 2007), and *network-based* lexical  
 74 resources: *WordNet* (Fellbaum 1998; Miller 1990), *MindNet* (Richardson et al.  
 75 1998), *HowNet* (Dong and Dong 2006), and *Pathfinder* (Schvaneveldt 1989). There  
 76 are Longman’s *Language Activator* (Summers 1993) and *OneLook*<sup>3</sup>, which, akin  
 77 to BabelNet (Navigli and Ponzetto 2012), combines a dictionary (WordNet) and an  
 78 encyclopedia (Wikipedia). Besides all this there are interesting proposals coming  
 79 from Fontenelle (1997), Sierra (2000), Moerdijk (2008), and Mel’çuk (Mel’çuk  
 80 and Polguère 2007). Finally, there is MEDAL (Rundell and Fox 2002), a thesaurus  
 81 produced with the help of Sketch Engine (Kilgarriff et al. 2004).

82 In parallel to dictionary making a lot of progress has been made by psycholinguists  
 83 who study the time course of lexical access (Levelt et al. 1999), word associations  
 84 (de Deyne and Storms 2015) and the structure, i.e. organization of the mental lexicon  
 85 (de Deyne et al. 2016).<sup>4</sup> Clearly, a lot has happened during the last two decades, yet  
 86 more can be done especially with respect to indexing (the organization of the data)  
 87 and navigation.

88 This paper is organized as follows. I start by providing evidence that storage does  
 89 not guarantee access. That this holds for humans has been shown already 50 years  
 90 ago (Tulving and Pearlstone 1966), in particular via Brown and McNeill’s (1966)  
 91 seminal work devoted to the *tip-of-the-tongue problem*. I will show here that this  
 92 can also hold for machines. The assumption that what is stored can also be accessed  
 93 (anytime), is simply wrong. To illustrate this claim I will compare an extended version  
 94 of WN (Mihalcea and Moldovan 2001) to an equivalent resource based on Wikipedia

<sup>2</sup>More or less well specified thoughts (concepts, elements of the word’s definition), or somehow related elements: collocations, i.e. associations (elephant: tusk, trunk, Africa).

<sup>3</sup><https://www.onelook.com>

<sup>4</sup>For a short survey of some of this work see Zock et al. (2009), and Zock (2015a, b).

95 (WiPe). Next, I will discuss under what conditions WN is adequate for word access,  
 96 and finally, I will sketch a roadmap describing the steps to be performed in order to  
 97 go beyond this very popular resource. The goal is to build a navigational tool (index,  
 98 association network) allowing authors to go from the word they know (word available  
 99 when being in the ToT state) to the word they are looking for (target). Before doing  
 100 so, I will present though my theory concerning the dialogue between the dictionary  
 101 user and the lexical resource.

## 102 14.2 Storage Does Not Guarantee Access

103 To test this claim let me describe here briefly an experiment carried out with a  
 104 colleague of mine (Zock and Schwab 2011). We ran a small experiment, comparing  
 105 an extended version of WN (henceforth, WN-x) and *Wikipedia*, which we converted  
 106 into a lexical resource. Our goal was not so much to check the quality of WN or  
 107 any of its extensions as to show, firstly, that storage does not guarantee access and,  
 108 secondly, that access depends on a number of factors like (a) quality of the resource  
 109 within which the search takes place (organisation, completeness), (b) index, and  
 110 (c) type of the query (proximity to the target).<sup>5</sup> Having two resources built with  
 111 different foci, our goal was to check the efficiency of each one of them with respect  
 112 to word access. For practical reasons we considered only direct neighbors. Hence,  
 113 we defined a function called *direct neighborhood*, which, once applied to a given  
 114 window (sentence/paragraph,<sup>6</sup> produces all its co-occurrences. Of course, what holds  
 115 for *direct associations* (our case here), holds also for indirectly related words, that  
 116 is, words whose distance >1 (mediated associations).

## 117 14.3 Comparisons of the Two Resources

118 Table 14.1 shows the results produced by WN-x and WiPe for the following, randomly  
 119 given inputs: ‘wine’, ‘harvest’ or their combination (‘wine + harvest’).

120 Our goal was to find the word ‘vintage’. As the results show, ‘harvest’ is a better  
 121 query term than ‘wine’ (488 vs. 30 hits), and their combination is better than either

<sup>5</sup>To show the relative efficiency of a query, D. Schwab has developed a website in Java as a servlet. Usage is quite straightforward: people add or delete a word from the current list, and the system produces some output. The output is an ordered list of words, whose order depends on the overall score [i.e. the number of co-occurrences between the input, i.e. ‘source word’ ( $S_w$ ) and the directly associated words, called ‘potential target word’ ( $PT_w$ )]. For example, if the  $S_w$  ‘bunch’ co-occurred five times with ‘wine’ and eight times with ‘harvest’, we would get an overall score or weight of 13: [(wine, harvest), bunch, 13]. Weights can be used for ranking (i.e. prioritizing words) and the selection of words to be presented, both of which may be desirable when the list becomes long.

<sup>6</sup>Optimal size is an empirical question, which may vary with the text type (encyclopedia vs. raw text).

**Table 14.1** Comparing two corpora with various inputs

Input	Output: WN-x	Output: WiPe
wine	<b>488 hits</b> grape, sweet, serve, France, small, fruit, dry, bottle, produce, red, bread, hold...	<b>3045 hits</b> name, lord characteristics, christian, grape, France, ... <u>vintage</u> (81st), ...
harvest	<b>30 hits</b> month, fish, grape, revolutionary, calendar, festival, butterfish, dollar, person, make, wine, first, ...	<b>4583 hits</b> agriculture, spirituality, liberate, production, producing, ..., <u>vintage</u> (112th), ...
wine + harvest	<b>6 hits</b> make, grape, fish, someone, commemorate, person, ...	<b>353 hits</b> grape, France, <u>vintage</u> (3rd), ...

122 of them (6 hits). What is more interesting though is the fact that none of these terms  
 123 allows us to access the target, eventhough it is contained in the database of WN-x,  
 124 which clearly supports our claim that storage does not guarantee access. Things are  
 125 quite different for an index built on the basis of information contained in WiPe. The  
 126 same input, ‘wine’ evokes many more words (3045 as opposed to 488, with ‘*vintage*’  
 127 in the 81st position). For ‘harvest’ we get 4583 hits instead of 30, ‘*vintage*’ occurring  
 128 in position 112. Combining the two yields 353 hits, which pushes the target word to  
 129 the third position, which is not bad at all. I hope that this example is clear enough  
 130 to convince the reader that it makes sense to use real text (ideally, a well-balanced  
 131 corpus) to extract from it the information needed (associations) in order to build an  
 132 index allowing users to find the elusive word.

133 One may wonder why we failed to access information contained in WN and why  
 134 WiPe performed so much better. We believe that the relative failure of WN is mainly  
 135 due to the following two facts: the size of the corpus (114,000 words as opposed to  
 136 3,550,000 for WiPe), and the number of syntagmatic links, both of which are fairly  
 137 small compared to WiPe. Obviously, being an encyclopedia, WiPe contains many  
 138 more *syntagmatic* links than WN. Of course, one could object that we did not use the  
 139 latest release of WN which contains many more words (147,278 words, clustered  
 140 into 117,659 synsets). True as it is, this would nevertheless not affect our line of  
 141 reasoning or our conclusion. Even in this larger resource we may fail to find what  
 142 we are looking for because of the lack of *syntagmatic links*.<sup>7</sup>

143 As mentioned already, the weak point is not so much the quantity of the data,  
 144 as the quality of the index (the relative sparsity of links). Yet, in order to be fair

<sup>7</sup>It should be noted though that serious efforts have been made to enrich WN by adding syntagmatic links (Bentivogli and Pianta 2004) and various kinds of encyclopedic information: topic signatures (Agirre et al. 2001), domain-specific information (Boyd-Graber et al. 2006; Gliozzo and Strapparava 2008; Fernando 2013), etc., but none of them seems to be integrated in the version accessible via the web interface (<http://wordnetweb.princeton.edu/perl/webwn>). Yet this is the one accessed by the ordinary language user who is generally either not able or willing to spend time to write an algorithm to integrate the different resources.

**Table 14.2** Comparing two corpora with various inputs

	Output: WN-x	Output: WiPe
ball	<b>346 hits</b> game, racket, player, court, volley, Wimbledon, championships, inflammation, ..., <u>tennis</u> (15th), ...	<b>4891 words</b> sport, league, football, hand, food, foot, win, run, game, ..., <u>tennis</u> (27th), ...
racket	<b>114 hits</b> break, headquarter, gangster, lieutenant, rival, kill, die, ambush, <u>tennis</u> (38th), ...	<b>2543 words</b> death, kill, illegal, business, corrupt, ..., <u>tennis</u> (72nd), ...
ball + racket	<b>11 hits</b> game, <u>tennis</u> , (2nd), ...	<b>528 hits</b> sport, strike, <u>tennis</u> (3rd), ...

145 towards WN, one must admit that, had we built our resource differently,—for exam-  
 146 ple, by including in the list of related terms, not only the directly evoked words,  
 147 i.e. potential target words, but all the words containing the source-word (wine) in  
 148 their definition (Bordeaux, Retsina, Tokay),—then we would get ‘vintage’, as the  
 149 term ‘wine’ is contained in its definition (‘vintage’: a season’s yield of ‘wine’ from  
 150 a vineyard). Note that in such cases even Google works often quite well, but see also  
 151 (Bilac et al. 2004; El-Kahlout and Oflazer 2004; Dutoit and Nugues 2002).

152 Last but not least, success may vary quite dramatically, depending on the input  
 153 (quality of the query). As you can see in Table 14.2, WN performs slightly better  
 154 than WiPe for the words ‘ball’, ‘racket’ and ‘tennis’. Yet, WiPe does not lag much  
 155 behind; additionally, it contains many other words possibly leading to the target  
 156 words (“player, racket, court”, ranked, respectively in position 12, 18 and 20).

157 Not being an encyclopedia, WN lacks most of them, though surprisingly, it con-  
 158 tains named entities like ‘Seles’ and ‘Graf’, two great female tennis players of the  
 159 past. Given the respective qualities of WN and WiPe one may well consider integrat-  
 160 ing the two by relying on a resource like *BabelNet* (Navigli and Ponzetto 2012).<sup>8</sup>  
 161 This could be done in the future. In the meantime let us take a closer look at WN and  
 162 its qualities with respect to word look up.

## 163 14.4 Under What Condition Is WN Really Good 164 for Consultation?

165 It is a well-known fact that WN is based on psycholinguistic principles (associations,  
 166 network, hierarchical structure, ...). What is less known though is the fact, that despite  
 167 its origins, it has never been built for consultation. It has been primarily conceived  
 168 for usage by machines: “WordNet is an online lexical database designed for use under

<sup>8</sup><http://lcl.uniroma1.it/babelnet/>.

169 program control.” (Miller 1995, p. 39). This being said, WN can nevertheless be used  
 170 for consultation, all the more as it is quite good at it under certain circumstances.

171 Remains the question under what conditions WN is able to reveal the elusive  
 172 target word. I believe that it can do so perfectly well provided that the following  
 173 three conditions are met:

- 174 (a) the *author knows* the *link* holding between the source word (input, say ‘dog’)  
 175 and the target, e.g. ([dog] + *synonym* = [? target]) → ([target = bitch]); ([dog] +  
 176 *hypernym* = [? target]) → ([target = canine]);  
 177 (b) the *input* (source word) and the *target* are *direct neighbors* in the resource. For  
 178 example, [seat]-[leg] (*meronym*); or [talk]-[whisper] (*troponym*), ...  
 179 (c) the *link* is *part* of WN’s database, e.g. ‘hyponym/hypernym’, ‘meronym’, ...

## 180 14.5 The Framework of a Navigational Tool 181 for the Dictionary of the Future

182 To access a word means basically to reduce the entire set of words stored in the  
 183 resource (lexicon), to one (target). Obviously, this kind of reduction should be per-  
 184 formed quickly and naturally, requiring as little time and effort (minimal number  
 185 of steps) as possible on the users’ side. Note that this process is knowledge based,  
 186 meaning that the user may have stored not only the elusive word but also other, some-  
 187 how related words. This is a very important point, as in case of failure, the dictionary  
 188 user may well start from any of these connected words.

189 When I wrote that WN is quite successful with regard to word look-up under  
 190 certain circumstances, I also implied that it is not so good when these conditions are  
 191 not met. More precisely, this is likely to occur when:

- 192 (a) the source (input) and the target are only *indirectly* related, the distance between  
 193 the two being greater than 1. This would be the case when the target (‘Steffi  
 194 Graf’) cannot be found directly in reponse to some input (‘tennis player’), but  
 195 only via an additional step, say, ‘tennis pro’—([tennis player] → [tennis pro])—  
 196 given as input at the next cycle, in which case it will at best only then reveal the  
 197 target.<sup>9</sup>  
 198 (b) the input (‘play’) and the target (‘tennis’) belong to different parts of speech  
 199 (see ‘tennis problem’, Fellbaum 1998);  
 200 (c) the prime and the target are linked via a *syntagmatic association* (‘smoke’-  
 201 ‘cigar’). Since the majority of relations used by WN connect words from the

<sup>9</sup>Note that the situation described is a potential problem for any association network. Note also that, even though Named Entities (NEs) are generally not contained in a lexicon, some of them have made it into WN. This is the case for some famous tennis players, like Steffi Graf. Anyhow, since NEs are also words, the point we are trying to make holds for both. Hence, both can be organized as networks, and whether access is direct or indirect depends on the relative proximity of the input (prime) with respect to the target word.



202 same part of speech, word access is difficult if the output (target) belongs to a  
 203 different part of speech than the input (prime)<sup>10</sup>;

204 (d) the user ignores the link, he cannot name it, or the link is not part of WN's repertory.  
 205 <sup>11</sup> Actually this holds true (at least) for nearly all syntagmatic associations.

206 Let us see how to go beyond this. To this end I present here briefly the principles of  
 207 the resource within which search takes place, as well as the required navigational aid  
 208 (categorical tree) to allow authors to find quickly the word they are looking for. Yet,  
 209 before doing so, let me clarify some differences between hierarchically structured  
 210 dictionaries and my approach.

211 While lexical ontologists (LO) try to integrate all words of a language into a neat  
 212 subsumption hierarchy, we try to group them mainly in terms of direct neighborhood.  
 213 More precisely, we try to build a lexical graph where all words are connected, regard-  
 214 less of whether we can name the link or not. Put differently, we try to build a hybrid  
 215 association network whose elements (words) are connected via typed and untyped  
 216 links. Both kinds of links are necessary for filtering, i.e. to ensure that the search  
 217 space is neither too big (typed links), nor too small (untyped links). Knowledge of  
 218 the relationship between the source and the target is an obvious asset, as it reduces  
 219 considerably the search space. Yet, untyped links are a necessary evil: they address  
 220 the fact that two words evoke each other. Hence, even if we cannot name the link, we  
 221 should still include the connected word in the list within which search takes place.  
 222 Otherwise, how can the user find it? Of course, untyped links can cause growth  
 223 of the search space. Yet, in order to avoid this problem we could group by category  
 224 the words devoid of a link (Fig. 14.1, step 2). Obviously, this approach yields a quite  
 225 different network than WN. Hence it will also produce different results than WN for  
 226 a given input (see Table 14.3).

227 Suppose we started from a broad term like 'food'. A lexical ontology like WN  
 228 would produce the entire list of objects referring to 'food' (hyponyms), while an  
 229 association network would only reveal typically evoked words {food, bread, noodles,  
 230 rice, fish, meat, cook, eat, buy, starving, good, expensive, fork, chopsticks....}. This  
 231 list contains, of course, a subset of the terms found in a LO (terms referring to  
 232 'food'), but it also contains syntagmatically related words (*origine*: France; *state*:  
 233 hungry, ...). Compare the respective results obtained by WN and the Edinburgh  
 234 Association Thesaurus.<sup>12</sup>

235 By taking a look at this second list one can see that it contains not only hyponyms,  
 236 that is, specific kinds of food (meat, cheese, ...), but also syntagmatically related  
 237 words (cook, good, France, ...), i.e. words typically co-occurring with the term  
 238 'food'. Note that our list may lack items like 'bagles', 'cheese' or 'olives'. This is  
 239 quite normal, if ever these words are not strongly associated with our input (food),  
 240 which does not imply, of course, that we cannot activate or find them. Had we given

<sup>10</sup>This being said, WN does have cross-POS relations, i.e. "morphosemantic" links holding among semantically similar words: observe (V), observant (Adj) observation (N).

<sup>11</sup>For example: 'well-known\_for', 'winner\_of', ...

<sup>12</sup><http://www.eat.rl.ac.uk> (see also: <http://rali.iro.umontreal.ca/word-associations/query/>).

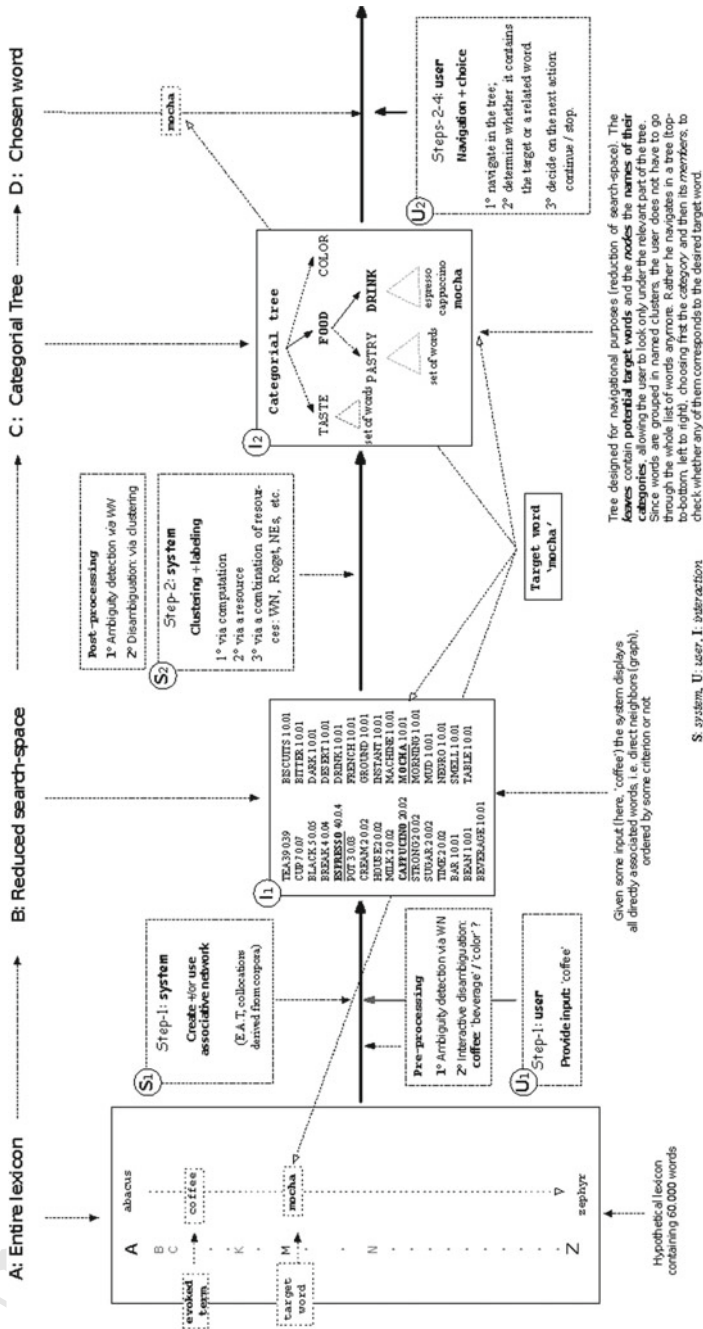


Fig. 14.1 Lexical access a four-step process for the user

**Table 14.3** The respective outputs produced by a *lexical ontology* (here WN) as opposed to an *association network*, here, the Edinburgh Association Thesaurus (E.A.T)

---

**WN:** *hypernym*: solid; *part\_holonym*: nutrient; **hyponyms**: leftovers, fresh\_food, convenience\_food, chocolate, baked\_goods, loaf, meat, pasta, health\_food, junk\_food, breakfast\_food, green\_goods, green\_groceries, coconut, coconut\_meat, dika\_bread, fish, seafood, butter, yoghurt, cheese, slop

---

**E.A.T:** at, drink, good, thought, dinner, eating, hunger, salad, again, apple, baby, bacon, bread, breakfast, case, cheese, consumption, cook, firm, fish, France, goo, great, hungry, indian, kitchen, lamb, loot, meal, meat, mix, mouth, noah, nosy, of, pig, please, poison, rotten, sausage, steak, stomach, storage, store, stuff, time, water, yoghurt, yum

---

241 ‘wine’ or ‘oil’ ‘green’ and ‘Greece’ as input, chances are that ‘cheese’ and ‘olives’  
 242 would pop up immediately, while they are burried deep down in the long list of food  
 243 produced by a LO.

244 Let us return to the problem of word access. Just as orientation in real world  
 245 requires tools (map, compass) we need something equivalent for locating a word in  
 246 a lexical resource. While the *semantic map* defines the territory within which search  
 247 takes place, the *lexical compass* guides the user, helping her to reach the goal (target  
 248 word). Obviously, the terms map and compass are but metaphors, as there are impor-  
 249 tant differences between world maps and lexical graphs (see below) on one hand,  
 250 and compasses sailors use and the tool an information seeker is relying on (human  
 251 brain) on the other. The map I have in mind is basically an association network. It  
 252 is a fully connected graph encoding all directly associated words given some input.  
 253 This kind of graph has many redundancies, and the links are not necessarily labeled.  
 254 In this respect it is very different from WN and even more so from the maps we  
 255 use when traveling in real world. Also, when using a world map the user generally  
 256 knows more or less precisely the destination or the relative location of the place he  
 257 is looking for, for example, south of Florence. He may also be able to deduce its  
 258 approximate location, eventhough she is not able to produce its name (Rome). This  
 259 does not hold in the case of a user resorting to a lexical resource (map) based on  
 260 associations. While the user may know the starting point (knowledge available when  
 261 trying to find the target, the elusive word), he cannot name the destination (target), as  
 262 if he could, there would be no search problem to begin with. The user is either able  
 263 to activate the word (in which case the problem is solved), or not. In this latter case  
 264 all he can do is to rely on available knowledge concerning the target, an assumption I  
 265 clearly make here. For example, users often know a related word, and they know how  
 266 it relates to the target: (part of the) meaning, sound, collocational, etc. Knowledge is  
 267 often fragmentary. Yet, incomplete as it may be, this kind of information may allow  
 268 us to help them to find the target, guiding him in a reduced, clearly marked search  
 269 space (details here below).

270 To get back to navigation in real world. In the case of spatial navigation it suffices  
 271 to know that ‘Rome’ is south of ‘Florence’, which is part of ‘Lazio’, and that it can be  
 272 reached by car in about 2 h. Having this kind of knowledge we could initiate search  
 273 in the area of ‘Lazio’, since ‘Lazio’ is an area south of ‘Tuscany’, the area containing

274 ‘Florence’. While this strategy works fine in the case of spatial navigation, it will not  
 275 work with lexical graphs. In this kind of network terms are related in many ways and  
 276 their strength may vary considerably. Hence, it is reasonable to show a term only if  
 277 it is above a certain threshold. For example, a term A (Espresso) being connected to  
 278 term B (coffee) may be shown only if it is sufficiently often evoked by B. Note that  
 279 eventhough words are organized in terms of neighborhood, the link between them  
 280 (explicited or not) may be of many other kinds than a spatial relation. In sum, the  
 281 links connecting words in an associative network are much more diverse than the  
 282 ones typically found in a lexical ontology.

283 As mentioned already, humans using world maps usually know the name of their  
 284 destination, whereas people being in the ToT state do not. Yet, even if they did,  
 285 they would not be able to locate it on the map. Lexical graphs are simply too big  
 286 to be shown entirely on a small screen.<sup>13</sup> In sum, we need a different approach:  
 287 search must be performed stepwise, taking place in a very confined space, composed  
 288 of the input and the direct neighbors (directly associated words). It is like a small  
 289 window moved by the user from one part of the graph to the next. If there are  
 290 differences between world maps and association networks (lexical graphs), there  
 291 are also important differences between a conventional compass and our navigational  
 292 tool. While the former automatically points to the north, letting the user compute  
 293 the path between his current location and the desired goal (destination, target), the  
 294 latter (brain) assumes the user to know, the goal, i.e. target word,<sup>14</sup> or its direction  
 295 (even if one does not know its precise location). While the user cannot name the  
 296 goal—he has only *passive knowledge* of it,—the system cannot guess it. However  
 297 it can make valuable suggestions. In other words, eventhough the system can only  
 298 make suggestions concerning the target or the directions to go (which word to use as  
 299 input for the next cycle), it is the user who finally decides whether the list contains  
 300 the target or not, and if so, in what direction to go. He is the only one to know which  
 301 suggestion corresponds best to the target (the word he has in mind) or which one of  
 302 them is the most closely connected to it. Of course, the user may go wrong, but as  
 303 experience shows his intuitions are generally quite good.

304 Before sketching a roadmap concerning the scenario of word access via the still-  
 305 to-be-built resource (association network), let me quickly provide some background  
 306 information concerning the users’ knowledge, a critical component in this kind of  
 307 dialogue.

---

<sup>13</sup> Associative networks contain many redundancies and are potentially endless, since they contain loops. For example, an input, say ‘Rome’ may well appear to be the direct neighbor of one of its outputs, ‘Italy’: ([Rome] → {[capital], [Italy], [city]}); ([Italy] → {[country], [France], [Rome]}).

<sup>14</sup> It has been shown over and over again that people being in the ToT state are able to identify immediately, and without making any mistakes the target word if it is shown to them, eventhough they could not name it. This is passive knowledge.

## 308 14.6 Navigation, a Fundamentally Cognitive Process

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

- 315 • *Declarative knowledge* is what we acquire when learning words (meaning, form,  
 316 spelling, usage), and this is the information generally encoded in dictionaries.  
 317 Obviously, in order to find a word or to find the information associated with it,  
 318 they must be stored, though this is not enough.
- 319 • Next, there is *meta-knowledge*, which also needs to be acquired. Being generally  
 320 unavailable for in(tro)spection, meta-knowledge reveals itself in various ways. For  
 321 example, via the information available when we *fail to access a word* (Schwartz  
 322 2006), or via the *query* we provide at the moment of launching a search. As word  
 323 association experiments have shown (Aitchison 2003) words always evoke some-  
 324 thing. Since this is true for all words one can conclude that all words are connected  
 325 in our mind, which implies that all words are accessible from anywhere like in a  
 326 fully connected graph.<sup>15</sup> All we have to do is to provide some input (source word,  
 327 available information) and follow then the path linking this input to the output  
 328 (target). Interestingly, people hardly ever start from words remotely related to the  
 329 target. Quite to the contrary, the words they give at the input (source words) tend  
 330 to be more or less direct neighbors of the target, requiring generally only one or  
 331 two steps for the solution, that is, they are hardly ever further away than the distance  
 332 of two (steps).<sup>16</sup> Also, dictionary users often know the type of relationship  
 333 holding between the input (prime) and the target, otherwise, why would lexicogra-  
 334 phers build thesauri, synonym- or collocation dictionaries? All these observations  
 335 lend support to our intuition that people have a considerable amount of (meta-)  
 336 knowledge concerning the organization of words in their mind, i.e. their mental  
 337 lexicon.

338 The notion of relationship has been nicely exploited by WN, which due to this  
 339 feature keeps the search space, i.e. a set of candidates among which the user has to  
 340 choose, quite small. The idea of relatedness has led lexicographers to build thesauri,  
 341 collocation- and synonym dictionaries. Obviously an input consisting only of a

---

<sup>15</sup>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.

<sup>16</sup>This is probably one of the reasons why we would feel estranged if someone provided as cue ‘computer’, while his target were ‘mocha’. The two are definitely not directly connected, though, there is a path between them, even though it is not obvious (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* → (*cappuccino*, *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).

342 simple word is hard to interpret. Is the user looking for a more general/specific  
 343 word, a synonym or antonym? Is the input semantically or phonetically related  
 344 to the target, or is it part of the target word’s definition (dog-animal)? In each  
 345 case the user is expecting a different word (or set of words) as output. Hence, in  
 346 order to enable a system to properly interpret the users’ goals we need this kind of  
 347 metalinguistic information (neighbor of the target, i.e. source word + relation to the  
 348 target) at the input.<sup>17</sup> If ever the user cannot provide it, the system is condemned  
 349 to make a rough guess, presenting all directly connected words. Obviously, such  
 350 a list can become quite large. This being so, it makes sense to provide the system  
 351 this kind of information to produce the right set of words, while keeping the search  
 352 space small.


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

362 The example here below illustrates to some extent these facts with regard to wordfind-  
 363 ing in an electronic resource. Suppose you are looking for a word conveying the idea  
 364 of a *large black-and-white herbivorous mammal of China*. Yet, for some reason you  
 365 fail to retrieve the intended form *panda*, even though you do know a lot concern-  
 366 ing the target. People being in this state, called the ToT-problem, would definitely  
 367 appreciate if the information they are able to access could be used to help them find  
 368 the target. Figure 14.1 illustrates the process of getting from a visual stimulus to  
 369 its corresponding linguistic output (word, expression) via a lexical resource. Given  
 370 an external stimulus (A) our brain activates a set of features (B) that ideally allow us  
 371 to retrieve the target form. If our brain fails, we use a fallback strategy and give part  
 372 of the activated information to a lexical resource (C) expecting it to filter its base  
 373 (D) in the hope to find the target (panda) or a somehow related word (E). As one  
 374 can see, we consider look-up basically as a two-step process. At step one the user  
 375 provides some input (current knowledge) to which the system answers with a set of  
 376 candidates, at step two the user scans this list to make her choice (Table 14.4).

---

<sup>17</sup>This has of course consequences with respect to the resource. To be able to satisfy the different user needs (goals, strategies) 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.

**Table 14.4** Lexical access a two-step process mediated by the brain and an external resource (lexicon)

A: Perceptual input, i.e. target	B: Associated features in the mental lexicon (brain)	C: Input to lexical resource	D: Lexical resource	E: Output of lexical resource
	<i>Type:</i> bear <i>Lives_in:</i> China <i>Features:</i> black patches <i>Diet:</i> eats bamboo	bear China	aardvark ... ... panda ... ... theorem ... zygote	<i>panda</i> polar bear

## 377 14.7 The Roadmap

378 Since alphabetically organized dictionaries are not very useful for onomasiological  
 379 search (language production), we follow WN by organizing words in terms of neigh-  
 380 borhood. All words are connected, and if possible, the links are given names. This  
 381 being so, we have a map supporting navigation. The user can enter the graph at any  
 382 point, to follow the links until having reached the target. Obviously, the kind of links,  
 383 as well as the presence/absence of link names (see below) are important with respect  
 384 to the search space (see below). A network devoid of link names yields many more  
 385 hits for a given input than a network containing the same set of words but whose  
 386 links are named. Imagine the number of possible outputs for ['tree'] compared to  
 387 [(‘tree’) + (‘synonym’/‘hypernym’)].

388 Let us now see quickly how to make all this work. Imagine an author wishing to  
 389 convey the name of a beverage commonly found in coffee shops (target: ‘mocha’).  
 390 Failing to do so, he reaches for a lexicon. Since dictionaries are too huge to be  
 391 scanned from cover (letter A) to cover (Z), I suggest to create a dialog between the  
 392 user and the computer to reduce incrementally the search space. The user provides  
 393 the input,<sup>18</sup>—word coming to his/her mind (source) when trying to access the tar-  
 394 get,—and the system produces as output a list of potential target words, in the hope to  
 395 enable the user to find the elusive word.

396 Note that, concerning the source and the target, there are basically three cases:  
 397 (A) the two are directly related, and the user knows their relationship; (B) the user  
 398 knows a direct neighbor, but he ignores the name of their relationship; (C) the source  
 399 and the target are not directly related, they are only indirect neighbors. Since the first  
 400 case (A) is quite well handled by WN, I will illustrate here only ‘B’, ignoring the  
 401 last case (C), as it can be solved indirectly by applying recursively the procedures

<sup>18</sup>This latter can be a single word—‘coffee’ in the case of target ‘mocha’—or a set of words, which in a normal communicative setting would yield a sentence, where the information seeker asks someone else to help him to find the elusive word.

402 proposed in ‘A’ and ‘B’.<sup>19</sup> To convey as simply as possible the rationale underlying  
 403 my approach let us make the following assumptions: (A) the user’s input is a  
 404 single word, here ‘coffee’ (step-1, Fig. 14.1); (B) the target, i.e. the elusive form,  
 405 is ‘mocha’; (C) the two are directly related in the resource in which search takes  
 406 places; (D) the nature of their relationship has not been specified by the user.<sup>20</sup>  
 407 Because of this last point—the relationship between the source and the target not  
 408 being given with the input—search space may grow considerably. To avoid this problem  
 409 I propose to present in a clustered and labeled form (categorical tree) all direct  
 410 associates of the input (step-2, Fig. 14.1).<sup>21</sup>

411 The user navigates in this tree, deciding on the category within which to look for  
 412 the target, and if he cannot find it in any of them, in what direction to go. If he could  
 413 find the target, search stops, otherwise the user will pick one of the associated terms  
 414 or provide an entirely new word, and the whole process iterates. The system will  
 415 then respond with a new set of proposals.

416 Two points, one concerning step-1, the other step-2. Ideally, the searchspace determined  
 417 at step-1 should contain the target word. As we have seen in the experiment  
 418 described here above (Sect. 14.3), WN failed to reveal a candidate though it was  
 419 stored in its database. This failure was due to a lack of syntagmatic associations,  
 420 that is, it was a side-effect of a design choice of how to connect words, or, which  
 421 words to connect. Hence, the search space proposed by WN in response to some  
 422 input was too small. At least it did not contain the target while its competitor (WiPe)  
 423 did.

424 The categories of our tree (step-2) resemble somehow those of Roget’s thesaurus.  
 425 Actually, they are not quite the same, but this is not really the point. What I would like  
 426 to stress is the fact that both the categories and the words composing the search-space  
 427 are dynamically computed in our case, while they are frozen, i.e. determined once  
 428 and for all in Roget. Hence, the set of words (search space) to be presented in the  
 429 categorical tree will be considerably smaller in our case than the ones displayed by  
 430 Roget. I believe that this kind of flexibility is a desirable feature as we cannot predict  
 431 the user’s input or his goals, we can only comply with them.

<sup>19</sup>This kind of wording can be generalized to a pattern for asking the following question: “What is the word for ‘[X] that [Y]?’”, where [X] is usually a hypernym and [Y] a stereotypical, possibly partial functional/relational/case description (action) of the target word. A similar pattern could be used for namefinding. For example, asking “What is the name of the <conqueror> of <empire>?” could yield ‘Pizarro’ or ‘Cortés’, depending on the value of the empire (Inca/Aztec). As one can see, the processes underlying wordfinding and namefinding are not very different.

<sup>20</sup>Note, that in order to determine properly the initial search space (step-1), we must have already well understood the input [mouse<sub>1</sub>/mouse<sub>2</sub> (rodent/device)], as otherwise our list will contain a lot of noise, presenting ‘cat, cheese’ together with ‘computer, mouse pad’ {cat, cheese, computer, mouse pad}, which is not quite what we want, since some of these candidates are irrelevant, i.e. beyond the scope of the user’s goal.

<sup>21</sup>This labeling is obligatory to allow for realistic navigation, as the list produced in response to the input may be very long and the words being of the same kind may be far apart from each other in the list. Hence it makes sense to structure words into groups by giving them appropriate (i.e. understandable) names so that the user, rather than looking up the entire list of words, searches only within a specific bag labeled by a category.



432 To get back to our roadmap. As one can see, the proposed method is quite straight-  
 433 forward, reducing considerably time and space needed for navigation and search.  
 434 Suppose that you had to locate a word in a resource of 50,000 words. If your input  
 435 triggered 100 direct associates, one of them being the target, then we would have  
 436 reduced in a single step the search space by 99.8%, limiting navigation and search to  
 437 a very small list. Suppose that our hundred words were evenly spread over 5 groups,  
 438 than search would consist in spotting the target in a list of 25 items: 5 being category  
 439 names and 20 being words within the chosen group.

440 A small note concerning the 2nd step. Step-2 yields a tree whose leaves are  
 441 *potential target words* and whose nodes are *categories*, which, while being also  
 442 words, are not at all the goal of the search. They are only the means to reach the goal.  
 443 Put differently, their function is orientational, guide the user during his search.

444 Words at the leave-level are potential target words, while the ones at the interme-  
 445 diate level (category names; preterminal nodes) are meant to reduce the number of  
 446 words among which to perform search, and to help the user to decide on the direction  
 447 to go. Hence, category names are reductionist and orientational (signposts), grouping  
 448 terminal nodes into a bag, signaling via their name not only the bag's content, but  
 449 also the direction to go. While the system knows the content of a bag, it is only the  
 450 user who can decide which of the bags is likely to contain the elusive word. Because,  
 451 eventhough he cannot name the target, he is the only one to know the target, be it  
 452 only passively and in fairly abstract terms. This is where the category names have  
 453 their role to play. In sum, it is not the system that decides on the direction to go next,  
 454 but the user. Seeing the names of the categories she can make reasonable guesses  
 455 concerning their content.

456 In conclusion, categories act somehow like signposts signaling the user the kind  
 457 of words he is likely to find choosing one bag rather than another. Indeed, knowing  
 458 the name of a category (fruit, animal), the user can guess the kind of words contained  
 459 in each bag (kiwi vs. crocodile). Assuming that the user knows the category of the  
 460 searched word,<sup>22</sup> she should be able to look in the right bag and take the best turn.  
 461 Navigating in a categorial tree, the user can search at a fairly high level (class) rather  
 462 than at the level of words (instances). This reduces not only the cognitive load, but  
 463 it increases also chances of finding the target, while speeding up search, i.e. the time  
 464 needed to find a word.

465 While step-1 is mainly a matter of *relatedness* ('wine' and 'red' being different in  
 466 nature, they are nevertheless somehow related), step-2 deals with *similarity*: there are  
 467 more commonalities between 'dogs' and 'cats' than between 'dogs' and 'trees'. Put  
 468 differently, the first two terms are more similar in kind than the last two. The solution  
 469 of the second step is certainly more of a challenge than the one of step-1 which  
 470 is largely solved, eventhough there is an issue of relevance: not all co-occurrences

---

<sup>22</sup>A fact which has been systematically observed for people being in the ToT state who may tell the listener that they are looking for the name of a "fruit typically found in a <PLACE>", say, New Zealand, in order to get 'kiwi'.

471 are really useful.<sup>23</sup> To put words into clusters is one thing, to give them names an  
 472 ordinary dictionary user can understand is quite another.<sup>24</sup> Yet, arguably building  
 473 this categorial tree is a crucial step, as it allows the user to navigate on this basis.  
 474 Of course, one could question the very need of labels, and perhaps this is not too  
 475 much of an issue if we have only say, 3–4 categories. We are nevertheless strongly  
 476 convinced that the problem is real, as soon as the number of categories (hence the  
 477 words to be classified) grows.

478 To conclude, I believe it is fair to say that the 1st stage seems to within reach, while  
 479 the automatic construction of the categorial tree remains a true challenge despite  
 480 some existing tools (word2vec) and the vast literature devoted to this specific or  
 481 otherwise strongly related problems (Zhang et al. 2012; Biemann 2012; Everitt et al.  
 482 2011).

483 One last point: to be truly useful, the user should provide as input not only a  
 484 word, but also a clue concerning the relationship between this input and his goal  
 485 (target word). Does he look for a semantically, formally (sound) or otherwise related  
 486 word with respect to the input? Since inputs can be interpreted in many ways, we  
 487 need additional information. Given some input, what is the user’s goal? Is he looking  
 488 for a synonym, hypernym or a similarly sounding word? Obviously, different goals  
 489 yield different searchspaces. This is a very important point. Authors searching for  
 490 a sound-related word to ‘right’ expect a different set of candidates (write, wright,  
 491 rite), from authors looking for its antonym (‘wrong’). WN takes advantage of this fact  
 492 eventhough only a subset of the ones mentioned here below are actually implemented.  
 493 Table 14.5 shows some possible links between some input (prime) and its directly  
 494 associated output (possible target).

495 Note that these links are of different sort: some are conceptual (1–13), others  
 496 are formal, i.e. they concern linguistic forms (15–19), and what we dubbed ‘free  
 497 association’ (14) concerns both. For more details concerning links or relations, take  
 498 a look at (Evens 2009; Green et al. 2002; Miller and Fellbaum 1992; Murphy 2003;  
 499 Nastase et al. 2013; Storjohann 2010).

## 500 14.8 Conclusion

501 Obviously, the human brain is a complex object and so is the process of accessing  
 502 words in the mental lexicon. My goal was not so much to address the problem of  
 503 complexity, i.e. the topology of the map of the mental lexicon. My goal was rather to  
 504 describe a method, that, once implemented, should help people to overcome the

---

<sup>23</sup>Take for example the Wikipedia page devoted to ‘panda’, and check which of the co-occurrences are those typically evoked when looking for this particular lexical concept.

<sup>24</sup>For example, while the sequence of hypernyms listed by WN for *horse* captures much of the phylogenetic detail a biologist would want to see recorded (horse → equine → odd-toed ungulate → ungulate → placental mammal → mammal → vertebrate → chordate → animal → organism → entity), most of these terms mean next to nothing to an ordinary dictionary user.

**Table 14.5** Possible links or associations between an input (cue/prime) and the target

	Type of relation	Description of the relation	Prime-target
1	<i>Hypernym</i>	A more <i>general</i> word	pie-pastry
2	<i>Hyponym</i>	A more <i>specific</i> word	fruit- <i>nut</i>
3a	<i>Meronym_substance</i>	A concept being a <i>substance</i> of another concept	<i>blood</i> -body
3b	<i>Meronym_part_of</i>	A concept being <i>part</i> of another concept	<i>ship</i> -fleet
3c	<i>Meronym_member_of</i>	A concept being a <i>member</i> of another concept	<i>kid</i> -family
4a	<i>Holonym_substance</i>	A concept having another concept as <i>substance</i>	sea- <i>salt</i>
4b	<i>Holonym_part_of</i>	A concept having another concept as <i>part</i>	tree- <i>leave</i>
4c	<i>Holonym_member_of</i>	A concept having another concept as <i>member</i>	team- <i>player</i>
5	<i>Cause to</i>	A verb expressing the <i>cause</i> of a result	kill- <i>die</i>
6	<i>Entailment</i>	A verb expressing an unavoidable <i>result</i>	buy- <i>have</i>
7	<i>Troponym</i>	A specific <i>way</i> to perform an action	drink- <i>sip</i>
8	<i>Part_of_meaning</i>	Part of the target word's definition	butter-milk
9	<i>Quality</i>	Typical quality, or inherent feature	snow- <i>cold</i>
10	<i>Co-occurrence</i>	Two concepts occurring frequently together	blue- <i>sky</i>
11	<i>Topically related</i>	Two concepts related by topic	sea-tide
12	<i>Used_for</i>	Instrumentally related words	fork-eating
13	<i>Made_of</i>	Substance or element used to make <object>	glass-sand
14	<i>Free association</i>	Can be any kind of link between two words	door-open
15	<i>Synonym</i>	Word expressing basically the <i>same</i> meaning	cup-mug
16	<i>Antonym</i>	A word meaning the <i>opposite</i>	dry-wet
17	<i>Sound (rhyme)</i>	Two similar sounding words	bad-mad/sad
18	<i>Homophones</i>	Words sounding alike, but spelled differently	right-write
19	<i>Anagrams</i>	Composed of same or similar components	cheater-teacher

505 ToT-problem. The method is radically knowledge-based, that is to say, it takes into  
 506 account knowledge users may have at the onset of consultation (see below).

507 I have started the paper by observing that word access remains a problem for  
 508 dictionary builders (Thumb 2004) and users alike, in particular for those being in the  
 509 production mode (Zock 2015a, b; Zock and Tesfaye 2015). Next I have shown that  
 510 word storage does not guarantee its access, even if the target is stored in a computer.  
 511 I have then analyzed some of the reasons why even a psycholinguistically motivated  
 512 resource like WN often fails to reveal the word authors are looking for.

513 Finally, I have recasted the problem within a cognitive framework, presenting a  
 514 roadmap of how to overcome the ToT-problem. The idea is to build a navigational  
 515 tool (hybrid association network) to help humans to find the word they are look-  
 516 ing for. The user provides the information coming to his mind when failing to access  
 517 the target word (input), and the resource produces a list of potential target words  
 518 (output). If the user can provide a (direct) neighbor of the target and its link, the

519 answer is generally straightforward. In the opposite case I suggest to present the  
 520 candidates in a labeled cluster-form (categorical tree) rather than as a huge, flat list.  
 521 While the system’s task is search-space reduction in step-1 (Fig. 14.1), its function  
 522 in step-2 (building of the categorical tree) is to support navigation. Just as it is unrea-  
 523 sonable to perform search in the entire lexicon, is it cumbersome to drill down huge  
 524 lists. This is why I suggested to cluster and label the outputs produced in response  
 525 to the query. After all, we want users to find the target quickly and naturally, rather  
 526 than drown them under a huge, unstructured (or poorly structured) list of words.

527 Note that there is at least one study supporting the idea that knowledge of link  
 528 names is useful. Nikolova et al. (2010) could show that word-finding is enhanced  
 529 when wordnets contain syntagmatic links. They describe a study where people struck  
 530 by aphasia used their resource, showing that retrieval was significantly better in this  
 531 case than when relying on a resource devoid of this information. They conceded  
 532 though that finding the first word to start communication with was still a problem. For  
 533 other related work see (Ferret 2015; Zock and Biemann 2016).

534 One last point: the success of the (yet-to-be-built) resource hinges critically on  
 535 three kinds of knowledge: (a) *factual knowledge*: to find a word it must exist, i.e. it  
 536 must be stored; (b) *metaknowledge*: to allow for word access, words must also be well-  
 537 organized, and the user must have some knowledge concerning this organization.  
 538 This amounts to knowing at least some of the words connected to the target, and the  
 539 relationship between some input (currently available word) and the goal (target word).  
 540 Put differently, in order to be able to provide a decent input (typically a relatively  
 541 close neighbor of the target), one must have at least a local view of the organization  
 542 of the mental lexicon; (c) *cognitive states*: they are revealed by the word(s) coming to  
 543 our mind when we search for a (word) form that we know, but cannot access. Solving  
 544 this problem is what this paper has been about. So far this is only a concept, but I hope  
 545 to be able to provide one evidence of its feasibility, as having such a tool would be  
 546 extremely precious for dictionary users being in the production mode.

## 547 References

- 548 Abrams, L., Trunk, D. L., & Margolin, S. J. (2007). Resolving tip-of-the-tongue states in young and  
 549 older adults: The role of phonology. In L. O. Randal (Ed.), *Aging and the elderly: Psychology,*  
 550 *sociology, and health* (pp. 1–41). Hauppauge, NY: Nova Science Publishers Inc.
- 551 Abel, A. (2012). Dictionary writing systems and beyond. In S. Granger, & M. Paquot (Eds.),  
 552 *Electronic lexicography* (pp. 83–106). Oxford: Oxford University Press.
- 553 Agirre, E., Ansa, O., Hovy, E., & Martinez, D. (2001). Enriching WordNet concepts with topic  
 554 signatures. Retrieved from <http://arxiv.org/abs/cs.CL/0109031>.
- 555 Aitchison, J. (2003). *Words in the mind: An introduction to the mental lexicon*. Oxford: Blackwell.
- 556 Benson, M., Benson, E., & Ilson, R. (2010). *The BBI combinatory dictionary of English*. Philadel-  
 557 phia: John Benjamins.
- 558 Bentivogli, L., & Pianta, E. (2004). Extending WordNet with syntagmatic information. In P. Sojka,  
 559 K. Pala, P. Smrz, C. Fellbaum, & P. Vossen (Eds.), *GlobalWor(l)dNet Conference, Proceedings*  
 560 (pp. 47–53). Brno: Masaryk University.

- 561 Benjamin, M. (2014). Collaboration in the production of a massively multilingual lexicon. In *LREC*  
 562 *Conference Proceedings* (pp. 211–215). Reykjavik.
- 563 Bernstein, T. (1975). *Bernstein's reverse dictionary*. New York: Crown.
- 564 Biemann, C. (2012). *Structure discovery in natural language*. Berlin: Springer.
- 565 Bilac, S., Watanabe, W., Hashimoto, T., Tokunaga, T., & Tanaka, H. (2004). Dictionary search based  
 566 on the target word description. In *Proceedings of the Tenth Annual Meeting of The Association*  
 567 *for Natural Language Processing* (pp. 556–559). Tokyo.
- 568 Boissière, P. (1862). *Dictionnaire analogique de la langue française: Répertoire complet des mots*  
 569 *par les idées et des idées par les mots*. Paris: Auguste Boyer.
- 570 Boyd-Graber, J., Fellbaum, C., Osherson, D., & Schapire, R. (2006). Adding dense, weighted,  
 571 connections to WordNet. In P. Sojka, Ks. Choi, C. Fellbaum, & P. Vossen (Eds.), *Proceedings of*  
 572 *the Global WordNet Conference 2006* (pp. 29–35). Brno: Masaryk University.
- 573 Brown, A. S. (1991). The tip of the tongue experience: A review and evaluation. *Psychological*  
 574 *Bulletin*, 10, 204–223.
- 575 Brown, R., & Mc Neill, D. (1966). The tip of the tongue phenomenon. *Journal of Verbal Learning*  
 576 *and Verbal Behaviour*, 5, 325–337.
- 577 Casares, J. (1942). *Diccionario ideológico de la lengua española*. Barcelona: Gustavo Gili.
- 578 de Deyne, S., & Storms, G. (2015). Word associations. In J. R. Taylor (Ed.), *The Oxford handbook*  
 579 *of the word*. Oxford, UK: Oxford University Press.
- 580 de Deyne, S., Verheyen, S., & Storms, G. (2016). Structure and organization of the mental lexicon: A  
 581 network approach derived from syntactic dependency relations and word associations. In *Towards*  
 582 *a theoretical framework for analyzing complex linguistic networks* (pp. 47–79). Berlin: Springer.
- 583 Dong, Z., & Dong, Q. (2006). *HowNet and the computation of meaning*. London: World Scientific.
- 584 Dornseiff, F., Wiegand, H. E., & Quasthoff, U. (2004). *Der deutsche Wortschatz nach Sachgruppen*.  
 585 Berlin/New York: de Gruyter.
- 586 Dutoit, D., & Nugues, P. (2002). A lexical network and an algorithm to find words from definitions.  
 587 In F. van Harmelen (Ed.), *Proceedings of the 15th European Conference on Artificial Intelligence*  
 588 (pp. 450–454). Amsterdam: IOS Press.
- 589 Edmonds, D. (Ed.). (1999). *The Oxford reverse dictionary*. Oxford: Oxford University Press.
- 590 El-Kahlout, I. D., & Ofłazer, K. (2004). Use of WordNet for retrieving words from their meanings.  
 591 In P. Sojka, K. Pala, P. Smrž, C. Fellbaum, & P. Vossen (Eds.), *Proceedings of the Global Wordnet*  
 592 *Conference* (pp. 118–123). Brno: Masaryk University.
- 593 Evens, M. W. (2009). *Relational models of the lexicon: Representing knowledge in semantic net-*  
 594 *works*. Cambridge: Cambridge University Press.
- 595 Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis*. Chichester: John Wiley &  
 596 Sons, Ltd.
- 597 Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database and some of its applications*.  
 598 Cambridge: MIT Press.
- 599 Fergusson, R., & Fergusson, R. (1985). *The Penguin rhyming dictionary*. London: Penguin.
- 600 Fernando, S. (2013). *Enriching lexical knowledge bases with encyclopedic relations* (Doctoral  
 601 dissertation). University of Sheffield.
- 602 Ferret, O. (2015). Typing relations in distributional thesauri. In N. Gala, R. Rapp & G. Bel-Enguix,  
 603 (Eds.), *Language Production, Cognition, and the Lexicon*. Springer, 113–134.
- 604 Fontenelle, T. (1997). *Turning a bilingual dictionary into a lexical-semantic database*. Tübingen:  
 605 Max Niemeyer.
- 606 Gliozzo, A., & Strapparava, C. (2008). *Semantic domains in computational linguistics*. Berlin:  
 607 Springer.
- 608 Granger, S., & Paquot, M. (Eds.). (2012). *Electronic lexicography*. Oxford: Oxford University Press.
- 609 Green, R., Bean, C. A., & Myaeng, S. H. (2002). *The semantics of relationships*. Dordrecht: Kluwer.
- 610 Hanks, P. (2012). Corpus evidence and electronic lexicography. In S. Granger & M. Paquot, (Eds.),  
 611 *Electronic lexicography* (pp. 57–82). Oxford: Oxford University Press.
- 612 Kahn, J. (1989). *Reader's Digest reverse dictionary*. London: Reader's Digest.

- 613 Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). *The sketch engine*. In G. Williams & S.  
 614 Vessier (Eds.), *Proceedings of the Eleventh EURALEX International Congress* (pp. 105–116).  
 615 Lorient, France: UBS.
- 616 Kilgarriff, A., & Kosem, I. (2012). Corpus tools for lexicographers. In S. Granger & M. Paquot  
 617 (Eds.), *Electronic Lexicography* (pp. 31–56). Oxford: Oxford University Press.
- 618 Levelt, W., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production.  
 619 *Behavioral and Brain Sciences*, 22, 1–75.
- 620 Levelt, W. J. (1993). *Speaking: From intention to articulation*. Cambridge: MIT Press.
- 621 Levelt, W. J. (2001). Spoken word production: A theory of lexical access. *Proceedings of the*  
 622 *National Academy of Sciences*, 98(23), 13464–13471.
- 623 Mel’čuk, I., & Polguère, A. (2007). *Lexique actif du français: l’apprentissage du vocabulaire fondé*  
 624 *sur 20 000 dérivations sémantiques et collocations du français*. Champs linguistiques. Bruxelles:  
 625 De Boeck.
- 626 Mihalcea, R., & Moldovan, D. (2001). Extended WordNet: Progress report. In *NAACL*  
 627 *2001—Workshop on WordNet and Other Lexical Resources* (pp. 95–100). Pittsburgh, USA.
- 628 Miller, G. (1991). *The science of words*. *Scientific American Library*. New York: W H Freeman &  
 629 Co.
- 630 Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, 38(11),  
 631 39–41.
- 632 Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet:  
 633 An on-line lexical database. *International Journal of Lexicography*, 3(4), 235–244.
- 634 Miller, G. A., & Fellbaum, C. (1992). Semantic networks of English. In B. Levin & S. Pinker (Eds.),  
 635 *Lexical and conceptual semantics* (pp. 197–229). Cambridge and Oxford, England: Blackwell.
- 636 Moerdijk, F. (2008). Frames and semagrams. Meaning description in the General Dutch Dictio-  
 637 nary. In *Proceedings of the Thirteenth Euralex International Congress* (pp. 561–570). Barcelona:  
 638 EURALEX.
- 639 Murphy, M. L. (2003). *Semantic relations and the lexicon: Antonymy, synonymy and other*  
 640 *paradigms*. Cambridge: Cambridge University Press.
- 641 Nastase, V., Nakov, P., Seaghdha, D. O., & Szpakowicz, S. (2013). Semantic relations between  
 642 nominals. *Synthesis Lectures on Human Language Technologies*, 6(1), 1–119.
- 643 Navigli, R., & Ponzetto, S. (2012). BabelNet: The automatic construction, evaluation and application  
 644 of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193, 217–250.
- 645 Nikolova, S., Tremaine, M., & Cook, P. R. (2010). Click on bake to get cookies: Guiding word-  
 646 finding with semantic associations. In *Proceedings of the 12th International. ACM SIGACCESS*  
 647 *Conference on Computers and Accessibility* (pp. 155–162). New York: ACM.
- 648 Rapp, B., & Goldrick, M. (2006). Speaking words: Contributions of cognitive neuropsychological  
 649 research. *Cognitive Neuropsychology*, 23(1), 39–73.
- 650 Richardson, S., Dolan, W., & Vanderwende, L. (1998). Mindnet: Acquiring and structuring semantic  
 651 information from text. In *Proceedings of the 17th international conference on Computational*  
 652 *linguistics, ACL-COLING ’98* (pp. 1098–1102). Montréal.
- 653 Robert, P., Rey, A., & Rey-Debove, J. (1993). *Dictionnaire alphabétique et analogique de la Langue*  
 654 *Française*. Paris: Le Robert.
- 655 Roget, P. (1852). *Thesaurus of English words and phrases*. London: Longman.
- 656 Rundell, M., & Fox, G. (Eds.). (2002). *Macmillan English dictionary for advanced learners*. Oxford:  
 657 Macmillan.
- 658 Schvaneveldt, R. (Ed.). (1989). *Pathfinder associative networks: Studies in knowledge organization*.  
 659 Norwood, New Jersey, US: Ablex.
- 660 Schwartz, B. L. (2002). *Tip-of-the-tongue states: Phenomenology, mechanism, and lexical*. Mah-  
 661 wah, NJ: Lawrence Erlbaum Associates.
- 662 Schwartz, B. L. (2006) Tip-of-the-tongue states as metacognition. *Metacognition and Learning*.  
 663 1(2), 149–158.
- 664 Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggensbach  
 665 (Ed.), *Perspectives on fluency* (pp. 200–219). Ann Arbor, MI: University of Michigan Press.

- 666 Sierra, G. (2000). The onomasiological dictionary: A gap in lexicography. In U. Heid, S. Evert,  
 667 E. Lehmann, & C. Rohrer (Eds.), *Proceedings of the Ninth Euralex International Congress*  
 668 (pp. 223–235). Stuttgart: IMS, Universität Stuttgart.
- 669 Storjohann, P. (Ed.). (2010). *Lexical-semantic relations: Theoretical and practical perspectives*.  
 670 Amsterdam: John Benjamins Publishing.
- 671 Summers, D. (1993). *Language Activator: The world's first production dictionary*. London: Long-  
 672 man.
- 673 Thumb, J. (2004). *Dictionary look-up strategies and the bilingualised learner's dictionary. A think-*  
 674 *aloud study*. Tübingen: Max Niemeyer Verlag.
- 675 Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for  
 676 words. *Journal of Verbal Learning and Verbal Behavior*, 5, 381–391.
- 677 van Sterkenburg, P. (2003). Onomasiological specifications and a concise history of onomasiolog-  
 678 ical dictionaries. In P. van Sterkenburg (Ed.), *A practical guide to lexicography* (pp. 127–143).  
 679 Amsterdam: John Benjamins Publishing.
- 680 Webster, M. (2007). Merriam Webster's rhyming dictionary. Merriam-Webster, Inc. Springfield,  
 681 Massachusetts.
- 682 Zhang, Z., Gentile, A., & Ciravegna, F. (2012). Recent advances in methods of lexical semantic  
 683 relatedness – A survey. *Journal of Natural Language Engineering*, 19(4), 411–479.
- 684 Zock, M. (2015a). 'Errare humanum est'. Refusing to 'appreciate' this fact could be a big *mistake!*  
 685 In G. Adda, M. Adda-Decker, J. Mariani, V. Barbu Mititelu, D. Tufis, & I. Vasilescu (Eds.),  
 686 *Errors by Humans and Machines in multimedia, multimodal and multilingual data processing*.  
 687 *Proceedings of ERRARE 2015*. Bucharest: Romanian Academy Publishing House.
- 688 Zock, M. (2015b). Introduction to the special issue of 'cognitive aspects of natural language pro-  
 689 cessing' (Words in books, computers and the human mind). *Journal of Cognitive Science*, 16(4),  
 690 355–378. Institute for Cognitive Science, Seoul National University ([http://j-cs.org/gnuboard/  
 691 bbs/board.php?bo\\_table=\\_\\_vol016i4](http://j-cs.org/gnuboard/bbs/board.php?bo_table=__vol016i4)).
- 692 Zock, M., & Biemann, C. (2016). Towards a resource based on users' knowledge to overcome the  
 693 Tip-of-the-Tongue problem. In *Proceedings of the COLING Workshop 'Cognitive Aspects of the  
 694 Lexicon' (CogALex-V)* (pp. 57–68) Osaka, Japan.
- 695 Zock, M., Ferret, O., & Schwab, D. (2010). Deliberate word access: An intuition, a roadmap and  
 696 some preliminary empirical results. *International Journal of Speech Technology*, 13(4), 107–117.
- 697 Zock, M., & Schwab, D. (2011). Storage does not guarantee access. The problem of organizing and  
 698 accessing words in a speaker's lexicon. *Journal of Cognitive Science*, 12(3), 233–258. Institute  
 699 for Cognitive Science, Seoul National University.
- 700 Zock, M., & Tesfaye, D. (2015). Automatic creation of a semantic network encoding *part\_of*  
 701 relations. *Journal of Cognitive Science*, 16(4), 431–491. Institute for Cognitive Science, Seoul  
 702 National University.
- 703 Zock, M., Wandmacher, T., & Ovchinnikova, E. (2009). Are vector-based approaches a feasible  
 704 solution to the « tip-of-the-tongue » problem? S. Granger & M. Paquot (Eds.), *eLexicography in  
 705 the 21st century: New challenges, new applications* (pp. 355–366). Louvain-la-Neuve.