

Michael Rundell and Nick Ham

A New Conceptual Map of English:

Abstract

The *Longman Language Activator* (published 1993) is a conceptually-organized dictionary that was designed and written, from scratch, specifically to meet the *encoding* needs of learners of English. The book's macrostructure represents a completely new conceptual map of English.

This paper describes how the conceptual framework was developed and how it was influenced by earlier models; how it works in practice to supply users' language-production needs; and what steps were taken to test and validate the system while it was under development.

0. Introduction

The *Longman Language Activator*, published in 1993, is subtitled 'the world's first production dictionary' because it was specifically designed and written to meet the *encoding* (as opposed to *decoding*) needs of intermediate to advanced learners of English. This rationale fundamentally affects the content and organisation of the book. One consequence, for example, is that entire classes of lexical items are systematically excluded – most notably what can broadly be described as 'real-world' nouns. Names of flora and fauna, machines and vehicles, or items of clothing and furniture are omitted on the grounds that they do not, generally speaking, pose any encoding problems that a good bilingual dictionary could not satisfactorily resolve. (And it is interesting to note that while nouns make up about 60% of all headwords in conventional dictionaries, they represent fewer than 25% of headwords in the *Activator*). The most striking feature of the book, however, is the fact that it is conceptually organized: although individual items *can* be accessed alphabetically, the dictionary's basic unit of organization is the 'Concept' or 'Key Word' entry, of which there are just over 1000.

The *Activator's* conceptual structure represents a completely new conceptual map of English. Its development is the outcome of a major corpus-based research project extending over 5 years, so this is clearly a very large topic. The present paper will restrict itself to dealing with some of the main issues of interest, focusing on the following areas:

1. antecedents: earlier conceptual systems, and their influence (if any) on the *Activator*
2. characteristics of the conceptual network
3. methodology: how the conceptual map was built up

4. **functionality:** how the system works in practice to take the dictionary user from (pre-lexical) meaning to an appropriate word or phrase to express this meaning
5. **validation:** methods that were used to ensure that the system actually delivered

1. Antecedents

While broadly belonging to the onomasiological (meaning-to-word) rather than semasiological (word-to-meaning) tradition, the *Activator* marks a radical departure from earlier models of conceptual organization. The best-known of these is of course *Roget's Thesaurus*, first published in 1852 but itself drawing on much earlier philosophically-motivated models connected with the search for a universal language.

Roget constructed a quasi-Linnaean taxonomy on 5 levels. At the top are the 6 major 'Classes' (with names such as 'Volition' and 'Affection'), below which are 'Divisions', then 'Sections' and then – the main organizing unit – the 990 'Heads', with names such as 'Tergiversation', 'Inutility', and 'Sufficiency'. Individual words (and sometimes phrases) are then grouped according to word-class within these Heads.

March's *Thesaurus-Dictionary* (March and March 1902) groups vocabulary items into typically bipolar conceptual sets such as 'AMITY-HOSTILITY' and 'FAVORITE-ANGER', and so to some extent prefigures Miller et al's *Wordnet*. A more recent attempt, Julius Laffal's *A Concept Dictionary of English* (Laffal 1973) establishes 118 major conceptual domains, with names such as LITL (covering vocabulary that refers to 'smallness, brevity, and diminution') and 'SUB' (with 'references to subordination, dependence, and subjection'). Laffal then tags each of 23,000 English words with an appropriate domain name, or in some cases with the names of two domains: thus **amputation** gets the (self-explanatory) tags BODY SEP, while **adultery** is intriguingly labelled BAD SEX. Laffal's motivation (he is a clinical psychologist) is to provide a tool for better understanding the anomalous speech of psychotic patients. The dictionary proceeds, in other words, from word to concept rather than vice versa and 'provides a means of looking through a speaker's language to the concepts which lie behind it'.

It will be immediately apparent that what all these systems have in common¹ is first, a 'top-down' approach, whereby a classification and/or taxonomy is constructed *a priori*, and individual lexical items are then assigned to relevant categories within the structure; and, secondly, a highly abstract character that severely limits their usefulness for non-fluent speakers.

In both these respects, the *Activator* takes a fundamentally different approach. Its conceptual structure is consciously atheoretical, and is

pedagogically – motivated rather than based on philosophical notions. And it was developed (see Section 3) in a heuristic, largely ‘bottom-up’ way.

2. Characteristics of the conceptual network

Probably the closest relation to an *Activator* concept is what cognitive psychologists, starting with Rosch, have called a ‘basic-level’ category or concept. The basic level is the home of general high-frequency terms that stand hierarchically speaking, between terms and subordinates: thus **chair** is a basic-level category with the superordinate **furniture** above it and a whole range of more specialized hyponyms below it (**deckchair, armchair, director’s chair**, and so on). The basic level is defined as being (among other things):

- * the level at which subjects are fastest at identifying category members
- * the first level named and understood by children
- * the level at which most of our knowledge is organized
(based on Lakoff 1987:46)

The literature here tends to focus on the kind of ‘real-world’ lexis that the *Activator* specifically excludes.² But the theory holds good for what might be called the language of predication – or words that say something *about* a person or thing as opposed to words that *name* the person or thing. Key words in the *Activator*, such as PROUD, ANGRY, and SAD/UNHAPPY, can legitimately be seen as basic-level concepts. Above them all is the theoretical superordinate EMOTIONS (theoretical because this level in the hierarchy has little pedagogical value and is therefore not used in the *Activator*). And below them is a wide range of subordinate vocabulary – not only less frequent near-synonyms that are semantically or connotationally more complex (such as **conceited, indignant, and despondent**) but also items connected to the key word through relationships such as causation and nominalization (such as **vanity, provoke, and depressing**).

The assumption underlying this conceptual organization is that ‘categorization begins at the basic level’ (Neisser 1987:22). The psychological processes involved in lexicalization – that is, in assigning a precise word or phrase to a given ‘entity’ (whether concrete or abstract) – are assumed to follow a fairly standard language-production model (e.g. Levelt 1989), whereby the speaker/writer first identifies a basic-level concept and from there maps his/her meaning onto word-forms in the mental lexicon that correspond to a specific meaning representation *within* the broad conceptual area.

One of the interesting philosophical questions that arises here (and it has important pedagogical implications too) is how far concepts of this type are language-and-culture-independent. Or to put it another way, is the *Activator’s* structure a conceptual map of English or a conceptual map of

language in general? There is obviously no simple answer to this, and it would be extremely rash to make any grand claims. Nevertheless, there is a reasonable consensus for the view that basic-level concepts are likely to be the most language-independent element in the lexicon. Superordinates may be culture-specific concepts, while more specialized lexis will often represent a cluster of semantic and/or pragmatic features that has no precise equivalent in another language.³ A system of categorization whose primary units are at the basic level seems to stand the best chance of being reasonably transparent to users whose first language is not English.

To conclude this section, it is worth saying a few words about what the *Activator's* concepts are *not*. These concepts should not be confused with the idea of semantic fields. A semantic-field approach to CRIME, for example, would include specific types of crime, such as murder, shoplifting, or rape. In the *Activator*, however, these notions are covered in the areas where they belong *conceptually*: murder at KILL (which is where a student wishing to encode this notion would naturally look), shoplift at STEAL, and rape at HAVE SEX.⁴

3. Methodology

The original plan for writing the *Activator* envisaged an initial stage in which the book's conceptual structure (the macrostructure) would be created. This would be followed by a much longer period in which the entries for individual headwords (the microstructure) would be researched and written. Things did not work out exactly as planned. The underlying goal of the process was always completely clear, namely, to provide the user – assumed to be an advanced learner of English, engaged in some form of language-production 'event' – with a simple and reliable system for finding the most appropriate lexical realization of the meaning s/he wishes to encode. But we were moving here in unknown territory, and the mechanisms for achieving this goal were by no means self-evident. In practice, the macrostructure continued to develop and change throughout the lifetime of the project. A reasonably robust version was in place at the end of the first two years, but the model was continually refined and was not *completely* finalized until close to the end of the project.

In outline, the process worked like this. The first stage was a systematic trawl of the vocabulary of a general – language pedagogical dictionary, the *Longman Dictionary of Contemporary English* (1987). A high percentage of *LDOCE* headwords did not survive this first pass, either because they denoted 'real-world' items outside the *Activator's* scope (such as **dog**, **elbow**, and **biscuit**),⁵ or because they were classed as 'reference-only' words which students might be expected to *recognize*, but would not be expected to produce (such as archaic, taboo, or highly formal language). Any word falling within the *Activator's* scope was then deconstructed in terms of its conceptual characteristics.

To take one example, **demonstrate** seems to have four main strands of meaning : the first relates to providing clear *proof* of something (*Galileo demonstrated that objects of different weights fall at the same speed*); the second to *showing* something or making it apparent (*anxious to demonstrate their concern for the homeless*); the third to explaining how something works or how to do something (*a ski instructor demonstrating turning techniques*); and the fourth, to taking part in some form of mass protest (*demonstrating against US involvement in the war*). This process was repeated for thousands of headwords, and in each case the question was asked : if a student wanted to encode this *meaning* but did not know the precise English *word* for it, where would s/he look to find it? Or to put it another way, what more general concept would s/he be likely to default to? For the first meaning of **demonstrate**, the default term would probably be PROVE, for the fourth meaning, PROTEST, and so on. PROVE and PROTEST thus became proto-concepts around which relevant lexical items gradually clustered. In a recursive process, the list of concepts and the list of headwords within each concept became progressively larger and more stable.

A typical concept might then look something like this:

concept name: ANGRY

lexical realizations of the concept: **angry, cross, enrage, fly off the handle, furious, incensed, mad, maddening, mollify, rage, stormy, wind someone up.**

The obvious need at this point was to take these undifferentiated lists and impose a structure that would enable users to navigate their way to the 'right' item without having to wade through a lot of irrelevant material. Detailed analysis tools were devised to make this process as systematic as possible. These included a number of key 'themes' according to which the basic concept might be structured, such as: intensification (very angry = **furious, livid**, etc); detensification (not very angry = **cross, annoyed**, etc); nominalization (**rage, anger**, etc); selectional restrictions (people can be **incensed**, meetings can be **stormy**, but not usually vice versa); causation (make someone angry = **enrage, wind up**, etc, things that make you angry are **infuriating, maddening**, etc); and 'anticausation' (stop someone being angry = **mollify, pacify**, etc). This is a very simplified, and to some extent idealized, version of a process that often seemed almost intractable. A great deal more could be said, but three points of special interest will be mentioned here.

First, we discovered that *Activator* concepts exhibited prototype effects in just the same way as more familiar basic-level concepts such as BIRD or VEGETABLE.⁶ Some category members were highly prototypical (and therefore easy to assign), while others were more marginal. Many lexical items gravitated very strongly to a particular concept, and there was a high level of agreement about this. To quote just a few examples among many:

chilly, freezing, and cool all belong unproblematically at COLD; **identical, uniform, and synonymous** at SAME; and **stride, saunter, and gait** at WALK.

Other words were weaker category members: **homesick**, for example, was fairly confidently assigned to SAD, but it is clearly a less prototypical member of the set than **miserable**. And for some words – thankfully a very small minority – there was no obvious conceptual home, and the eventual assignment entailed a fairly arbitrary choice between three or four possible locations.

The second point to mention here concerns the *names* given to the concepts. As with all aspects of the *Activator's* design, the book's pedagogical function crucially influenced this process. Section 1 (above) referred to the titles given by Roget to his main 'Heads', such as Tergiversation, Inutility, and Sufficiency. The corresponding conceptual groupings in the *Activator* are named CHANGE YOUR MIND, USELESS, and ENOUGH. In every case, the objective is to avoid abstraction and to use familiar, high frequency terms that combine accessibility and face-validity.⁷

A third point of interest concerns the place of multiword phrases in the *Activator*. Given that the dictionary's function is to supply appropriate lexicalizations of specific meanings, it follows that the purely *formal* distinction between words and phrases has no practical relevance. A fluent speaker wanting to convey the idea that a piece of machinery is not functioning properly is just as likely (in some contexts more likely) to say 'there's something wrong with it' or 'it's on the blink' as to resort to single-word equivalents such as **defective** or **faulty**. Consequently, sections in the dictionary covering a particular meaning area include *any* relevant lexical item, regardless of whether it is a word or phrase.

Almost 40% of *Activator* headwords are multiword lexemes. A high proportion of these are the type of ready-assembled, usually transparent, 'prefabs' which native-speakers routinely use to encode common notions. The concept for FOOD, for example, includes not only items such as **grub** and **refreshments**, but also the transparent (but not easily predictable) phrase **something to eat** ('Would you like something to eat?' 'Let's go out for something to eat' etc).⁸ Phrases of this type are highly typical of native-speaker performance, but in many cases have never before been covered in monolingual or bilingual dictionaries.⁹

Decisions about the assignment of phrasal items to concepts were critically affected by the environment in which a word most typically occurs in text. A word like **averse**, for example, is not shown at concepts such as DISLIKE or UNWILLING, because it nearly always occurs in the string **not be averse to** – and it is the meaning of the *whole unit* that determines its place in the conceptual structure. This approach finds some support in the psycholinguistic literature: there is evidence to suppose that many multiword items are both stored and retrieved as complete strings, rather than being assembled at the point of articulation.

4. Functionality : how the system works

The *Activator's* macrostructure is designed to take the user from a 'pre-verbal' message to an appropriate basic-level concept, and from there to a specific semantic area that broadly expresses the "right" meaning. This is achieved by means of a series of 'filters' that guide the user to a section in the book where a (manageable) range of options is offered. These are near-synonyms, and their carefully disambiguated definitions enable the user to select the *mot juste*. The underlying premise here is that when students want to convey a message which they lack the lexical resources to express precisely, they tend to start from the simple basic-level terms they already know. This resort to high-frequency default terms is a classic 'communication strategy' of the type described by Pit Corder and others (see e.g. Pit Corder 1983). And the use of a 'superordinate-plus-paraphrase' strategy (for example, 'steal from a shop' for **shoplift**, or 'listen in secret' for **eavesdrop**) is a pervasive feature of learners' text particularly at intermediate level and above.

In this sense, the *Activator's* navigation system incorporates strategies which most students already use, and enables them to make quite sophisticated lexical choices on the basis of vocabulary that they already know. At the same time, the encoding process used here can be seen as roughly analogous to the process by which fluent speakers access specific items from their mental lexicons.

5. Validation : testing the conceptual system

While the *Activator's* conceptual system was still under development, the model was tested and improved by means of both corpus evidence and user-research.

The Longman Learner's Corpus – a large collection of texts written by students of English – provided abundant evidence of the type of communication strategy described in the previous section. The frequency characteristics of lexical sets in the learners' corpus are strikingly different from those found in native-speaker corpora such as the Longman Lancaster Corpus. To take a simple example, in a set of words such as **interesting**, **fascinating**, **intriguing**, and **riveting**, the first item is easily the most frequent of the four in all types of text. But the preference for **interesting** is very much more marked in learners' text (at *all* levels of proficiency), and a similar pattern can be found in hundreds of similar sets.

The learners' corpus thus provided useful evidence of the type of default terms regularly favoured by students, and this helped to inform the process of concept creation and concept naming. It was found, too, that some types of lexical error suggested areas where learners' conceptual framework differed significantly from that of native-speakers: in these cases, too, the *Activator's* navigational systems were refined to pre-empt 'dead-end'

searches. Students from some language backgrounds, for example, showed some confusion between the concepts WAIT and EXPECT, so pointers were added to these Key Words in the *Activator* in order to ensure users would not waste time searching the 'wrong' entries.

Meanwhile a number of tests were carried out with students from different language backgrounds (e.g. Japanese, German, Romance languages) to investigate their communication strategies in situations where a precise target word was not in their active vocabulary. Students were given a text in their own language with certain words and phrases underlined. They were then asked to say how they would express these meanings in English. (A pre-test had already established that the subjects were *unlikely* to know the precise English equivalents for the test items). We were particularly interested here in validating concept names and, especially, in improving the names and configurations of concepts that seemed to us to lack face-validity. One problematical 'proto-concept' (which did not survive the development period) was named IN HARMONY, and included words such as **correspond** and **tally with**. The concept was reasonably coherent and may have worked quite well in a native-speaker version of the book, but we had serious doubts about its pedagogical value. The tests therefore included headwords from this grouping, and the results convinced us that the concept should be disbanded, with most of the material going to an expanded version of the concept SAME.

In parallel with these tests, a substantial piloting programme tested the effectiveness of *Activator* entries with students engaged in various language production tasks. The whole programme of user-research supporting the *Activator* should probably be the subject of a separate paper (its impact on the book's *microstructure*, for example, was considerable). But the main point to make here is that it had significant influence on the refinement of the dictionary's conceptual structure.

6. Conclusions

The conceptual framework created for the *Activator* is designed to provide fast and reliable access to groups of near-synonyms. Thereafter, the microstructure takes over, and the definitions for individual entries supply 'horizontal disambiguation' within each lexical set, enabling appropriate lexical choices to be made. It will be some time before we can judge how well the system actually works (though early reports are favourable), and it would be rash to imagine that it cannot be further improved. But we believe it to be a reasonably robust conceptual map, supported by extensive corpus research and informant-testing, and founded above all on strong pedagogical principles.

Notes:

- 1 The same point applies to *Hartrampf's Vocabulary Builder* (Gustavus A Hartrampf, Psychology Publishing Company: Marple, Cheshire 1929).
- 2 For example Rosch (1975), Cruse (1986: Chapter 6).
- 3 See e.g. DiMarco and Hirst (1993:34), who assume that "coarse denotational differentiation occurs at the language-independent conceptual level, and connotational and fine denotational differentiation occurs at the language-dependent level".
- 4 Longman has in fact already published a semantic-field dictionary which complements the *Activator's* conceptual orientation: the *Longman Lexicon of Contemporary English* (Tom McArthur, 1981) has a series of adjacent sections (C220 – C263) that deal specifically with the vocabulary of crime, punishment, and law enforcement.
- 5 But the *Activator* does include items such as **dog-eat-dog**, **dog-eared**, **elbow** (someone aside etc), and **take the biscuit**.
- 6 See e.g. Medin and Wattenmaker (1987 : 29) : "Concepts are represented in terms of properties that are only characteristic or probable of class members. Membership in a category can thus be graded rather than all-or-none, where the better members have more characteristic properties than the poorer ones".
- 7 See also Scholfield (1993).
- 8 Another interesting class of phrases start with the word **not**: for example, corpus evidence shows that a very typical way of saying someone is "stupid" is to say something like "He's not very bright" or "She's not all that smart". Where these tendencies were found to be frequent, we felt it was only logical to include such items in the range of options for expressing the concept; so the expressions mentioned here can be found at the *Activator* Key Word: STUPID/NOT INTELLIGENT.
- 9 As one might expect (e.g. from Pawley and Syder 1983), such phrases are particularly frequent in the spoken mode. The spoken component of the British National Corpus (10 million words, over half of it ordinary face-to-face conversation) was especially revealing in this respect.

REFERENCES

- Cruse, D.A. 1986 *Lexical Semantics*, Cambridge: Cambridge University Press.
- DiMarco, Chrysanne and Graeme Hirst 1993. "Usage notes as the basis for a representation of near-synonymy for lexical choice" in *Making Sense of Words* (Proceedings of the 9th Annual Conference of the University of Waterloo Centre for the new OED): 33–43
- Laffal, Julius 1973. *A Concept Dictionary of English*. Connecticut: Gallery Press.
- Lakoff, George 1987. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.
- Levelt, W.J.M. 1989. *Speaking : from Intention to Articulation*. Cambridge Mass: MIT Press.
- March, Francis Andrew and Francis A March Jr 1902. *March's Thesaurus – Dictionary*.
- Medin, Douglas L. and William D. Wattenmaker 1987. "Category cohesiveness, theories, and cognitive archaeology" in Ulric Neisser (ed) "*Concepts and Conceptual Development*". Cambridge: Cambridge University Press.
- Neisser, Ulric 1987. "From direct perception to conceptual structure" in Ulric Neisser (ed) *Concepts and Conceptual Development*. Cambridge : Cambridge University Press.
- Pawley, Andrew and Frances Syder 1983. "Two Puzzles for linguistic theory" in Jack C. Richards and Richard W. Schmidt (eds) *Language and Communication*. Harlow, Longman.
- Pit Corder, S. 1983. "Strategies of Communication" in Claus Faerch and Gabriele Kasper (eds) *Strategies in Interlanguage Communications*. Harlow : Longman.
- Rosch, E 1975. "Cognitive representations of semantic categories" in *Journal of Experimental Psychology* 104 : 192–233.
- Scholfield, Philip 1993. "The Conceptual map of English and the *Activator*" in *Longman Language Activator* F17–F19. Harlow : Longman.