Available online at www.sciencedirect.com
ScienceDirect
Cognitive Systems Research 9 (2008) 214–228
Cognitive Systems RESEARCH
www.elsevier.com/locate/cogsys

Economically organized hierarchies in WordNet and the Oxford English Dictionary

Action editor: Risto Miikkulainen

Mark A. Changizi

Department of Cognitive Science, Rensselaer Polytechnic Institute, Troy, NY 12180, United States

Received 29 January 2007; accepted 5 February 2008
Available online 5 March 2008

Abstract

Good definitions consist of words that are more basic than the defined word. There are, however, many ways of satisfying this desideratum. For example, at one extreme, there could be a small set of atomic words that are used to define all other words; i.e., there would be just two hierarchical levels. Alternatively, there could be many hierarchical levels, where a small set of atomic words is used to define a larger set of words, and these are, in turn, used to define the next hierarchically higher set of words, and so on to the top-level of very specific, complex words. Importantly, some possible organizations are more economical than others in the amount of space required to record all the definitions. Here I ask, How economical are dictionaries? I present a simple model for an optimal set of definitions, predicting on the order of seven hierarchical levels. I test the model via measurements from WordNet and the Oxford English Dictionary, and find that the organization of each possesses the signature features expected for an economical dictionary.

Keywords: Vocabulary; Hierarchy; Optimality; WordNet; Definition; Dictionary; Number of levels

1. Introduction

There are many ways to define a set of words using a small set of atomic words. On the one hand, each word could be defined directly in terms of atomic words, in which case there would just be two hierarchical levels to the "definition network": the bottom level set of atomic words, and the upper level (Fig. 1a). On the other hand, the small set of atomic words could be used to first define an intermediate level of words, and these words used, in turn, to define the target set of words; in this case there would be three hierarchical levels (Fig. 1b). Multiple intermediate levels are clearly possible as well. Depending on the sizes of the set of atomic words and the set of target words, some of these hierarchical organizations—e.g., some number of levels—will be more economical than others in the total amount of space required to write all the dictionary definitions (for Fig. 1, the "dictionary" in 1b is more economical). The question I take up in this paper is, *Are actual dictionaries economically organized?* And one central subquestion will be, Is the number of hierarchical levels in the dictionary consistent with an economical organization? As we will see, to a first approximation, dictionaries like WordNet and the Oxford English Dictionary do appear to have the signature features of one that is economically organized.

2. Signature features of an economically organized dictionary

The model optimal dictionary is assumed to be reductionistic, where each word is defined via words more basic (or less concrete, less specific, less complex) than itself. And, in addition, reductionistic dictionaries possess (as in

E-mail address: changizi@rpi.edu
URL: http://www.changizi.com

1389-0417/$ - see front matter
doi:10.1016/j.cogsys.2008.02.001

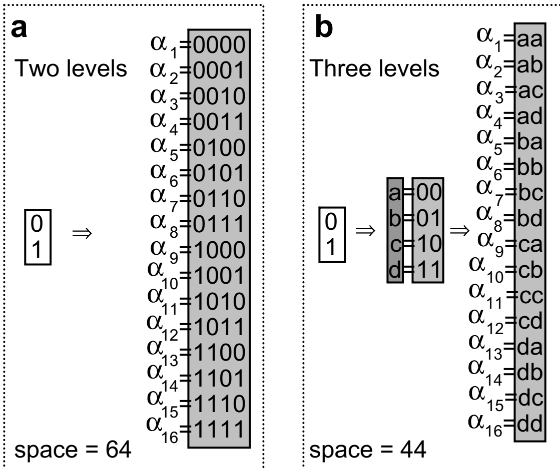

Fig. 1. Illustration that adding an intermediate level can decrease the overall space required to define a set of words. In each case 16 words must be given definitions (α_1 to α_{16}) via two atomic words (0 and 1). (a) The 16 words are directly defined via the two atoms, making for two hierarchical levels. Because each of the 16 target words requires definitions of (at least) length 4, the total number of words across all the "dictionary" definitions is $4 \times 16 = 64$ (and there are 18 words in the "dictionary"). (b) Four intermediate level words (a, b, c, and d) are defined from the atoms first, and these, in turn, are used to define the 16 target words, making three hierarchical levels. Each of the 16 definitions is half the lengh as before, but there are now four new definitions. The total number of words across all the "dictionary" definitions is now $2 \times 4 + 2 \times 16 = 40$ (and there are now 22 words in the "dictionary"). This "dictionary" is more economical (in total size) than the one in (a).

the toy example of Fig. 1) a set of bottom-level, or most-basic, atomic words who do *not* get their meaning by reference to other words. There are many ways of building a reductionistic dictionary, and my model will be the one that minimizes the overall amount of space needed to define all the words in the dictionary starting from the set of bottom-level, or atomic, words.

Before describing the prediction for an optimal dictionary, we need an empirical estimate of the number of bottom-level (or atomic) words actually in the dictionary, D_0 (the analogy of 0 and 1 in Fig. 1), and also an estimate of the total number of target words in the dictionary, D_{top} (the analogy of α_1 to α_{16} in Fig. 1). As one estimate of the total number of atomic words in English we use the number of words in WordNet for which there are no hypernyms (see also Section 3). When B is a kind of C, C is a hypernym of B. So, words without hypernyms are, in a sense, the most fundamental. In WordNet there are 10 such words (see Appendix B.1; see also Fellbaum, 1998). As a second estimate of the number of atomic words in English there are estimates from Wierzbicka and Goddard (from the natural semantics metalanguage approach to semantic analysis) who have provided evidence that there are approximately 60 words, called *semantic primes*, that cannot be further defined by simpler words (Goddard, 2006; Goddard & Wierzbicka, 2002; Wierzbicka, 1996). Accordingly, I will use the range of $D_0 \approx 10$–60 as a plausible range for the number of bottom-level words. Because my prediction will only be an order-of-magnitude one, for the total number of top-level, target, words in the dictionary I will for simplicity assume that it is on the order of $D_{top} \approx 10^5$ (lower than 141,755, which is the total number of nouns in WordNet). As we will see, the predicted signature features (including the predicted number of hierarchical levels) change little if these D_0 and D_{top} estimates change, say, by factors of 2 up or down.

Now we can make the optimality question more specific. What is the optimal way of defining $D_{top} \approx 10^5$ many words using $D_0 \approx 10$–60 fundamental words? Consider that if 10 fundamental words are used to define 10^5 words without any intermediate levels, this would require five word tokens per definition (ignoring redundancies), for a total space requirement of 500,000 word tokens, analogous to Fig. 1a. If, instead, $10^{5^{1/2}} = 172$ intermediate-level word types are first defined, and these, in turn, are used to define the 10^5 target word types, analogous to Fig. 1b, then the total amount of space needed for the definitions drops to a little over 223,992 word tokens, or less than half of what it was without the intermediate level. [To define $10^{5^{1/2}} = 172$ intermediate-level words via the 10 atomic words requires an average definition length of $\log(10^{5^{1/2}})/\log(10) = 5^{1/2} = 2.236$, for a total space for intermediate-level definitions of $10^{5^{1/2}} * (5^{1/2}) = 385$. These $10^{5^{1/2}} = 172$ words can then be utilized to define the 10^5 target words, and the average definition length of each of these is $\log(10^5)/\log(10^{5^{1/2}}) = 5^{1/2} = 2.236$ (the same length as the intermediate-level definitions, by construction), for a total space for top-level definitions of $10^5 * (5^{1/2}) = 223{,}607$. The total space for intermediate and top-level definitions is then $385 + 223{,}607 = 223{,}992$. Including the statement of the intermediate-level words themselves only adds a negligible 172 words to the sum. One can see that with only an extra space of 385 for the intermediate-level definitions—and perhaps a space 557 if one includes the intermediate word labels themselves—the dictionary size is reduced to 44.8% of its size when there was no intermediate level.]

More generally, Fig. 2a shows how the size of the set of all the definitions depends on the number of hierarchical levels, and the minimum occurs when there are seven levels, requiring about 150,000 word tokens across all the definitions, or a dictionary (including definitions) that is approximately 30% the size of the dictionary when there were only two levels. Dictionaries with 5 through 10 levels are all within 10% of optimal. These estimates are for the case of $D_0 = 10$. For $D_0 = 60$, the optimal number of levels is 5, and levels 4 through 6 are within 10% of optimal. These conclusions change little if the number of top-level words varies by a factor of two in either direction, as shown in Fig. 2b. Therefore, if the actual dictionary's organization is near optimal (i.e., within 10%), then there should be about 4–10 levels. For perfect optimality we would expect from 5 to 7 levels, as indicated by the highlighted band on the *y*-axis of Fig. 2b.

A second prediction follows from the fact that when there are more levels in the hierarchy, the growth in the

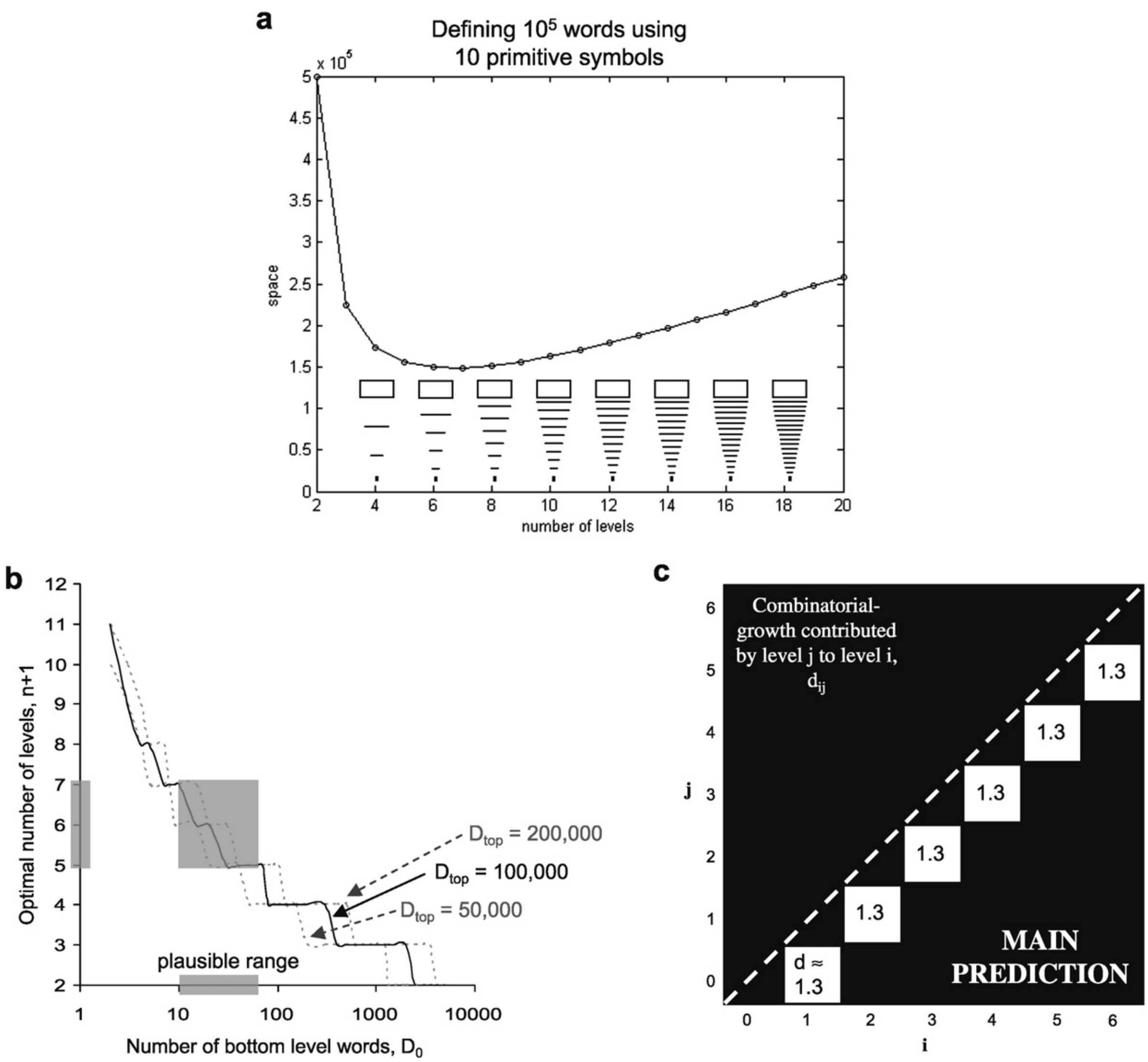

Fig. 2. (a) Total space required to define a lexicon ($D_{\rm top} = 10^5$ words with $D_0 = 10$ bottom-level words) versus the number of hierarchical levels. Shown in the bottom half of the plot are symbolic indicators of the hierarchical organization, showing the bottom level (black dot) and top-level (white bar), with variable numbers of intermediate levels in between (with the number of words per level rising for higher levels, indicated by the greater-width line segments). (b) Optimal number of hierarchical levels versus number of bottom-level words (D_0). The highlighted region along the x-axis shows the a priori plausible range for the number of bottom-level words, D_0 (namely, from 10 to 60). The highlighted region along the y-axis shows the consequent plausible range for the predicted (optimal) number of hierarchical levels, and varies only from 5 to 7 despite the plausible values for the x-axis ranging over nearly an order of magnitude. (c) The central prediction for an economically organized lexicon, showing how much each hierarchical level combinatorially contributes to define the words of every other level. The three signature features of the prediction are illustrated: (1) roughly seven levels (more weakly, about 4–10, see text), shown by the fact that the matrix is 7 by 7, (2) combinatorial growth from one level to the next that is roughly 1.3 (more weakly, from about 1.2 to 1.5, see text) [i.e., if D_i is the number of words of level i, then they are combinatorially employed to define $D_{i+1} = D_i^{1.3}$ many words of level $i+1$], and (3) each level contributes (via definitions) to the growth of the level just above it (i.e., a strict hierarchy), which is seen here by the contributions to the matrix being one below the diagonal, meaning level j contributes only to level $i = j+1$. The empirical test of this prediction may be seen in Fig. 5.

number of words from any level to the next is smaller (as illustrated by the insets in Fig. 2a), and this can be quantified by the exponent relating their sizes, which I call the *level–level combinatorial growth exponent*, d (Changizi, 2001, 2003). Defining $D_{\rm top}$ many words from D_0 many bottom-level words means that the dictionary has a *total combinatorial growth exponent* of $d_{\rm tot}$, where $D_{\rm top} = D_0^{d_{\rm tot}}$, and so $d_{\rm tot} = (\log D_{\rm top})/(\log D_0)$. If there are no intermediate levels in the hierarchy, then the level–level combinatorial growth exponent, d, is just the same as $d_{\rm tot}$. If there is one intermediate level, or three levels in all, then $D_{\rm top} = D_1^d = (D_0^d)^d = D_0^{d^2}$. It follows that $d = [(\log D_{\rm top})/(\log D_0)]^{1/2} = d_{\rm tot}^{1/2}$. More generally, if there are $n+1$ levels in the hierarchy (including the top and bottom), then the level–level combinatorial growth exponent is $d = d_{\rm tot}^{1/n}$. For $D_0 = 10$ and $D_{\rm top} = 10^5$, $d = d_{\rm tot}^{1/n} = [(\log D_{\rm top})/(\log D_0)]^{1/n} = [(\log 10^5)/(\log 10)]^{1/n} = 5^{1/n}$, and given that the optimum number of levels was $n+1=7$, $d=5^{1/(7-1)}=1.31$. Given that, for $D_0=10$, having from 5 to 10 levels is within 10% of optimal, this range of levels corresponds to a range of 1.5 down to 1.2 for the level–level combinatorial growth. If $D_0=60$ instead, then having 4–6 levels was near-optimal, and this corresponds to a level–level combinatorial growth range from 1.41 to 1.22. Therefore, if the actual dictionary's organization is near-optimal (i.e., within 10%), then the level–level combinatorial growth exponent should range from 1.5 to 1.2. For perfect optimality we would expect a level–level growth exponent of about 1.3.

A third feature of the model economically organized dictionary is that any given hierarchical level should contribute to the definitions of words in the level just one above it in the hierarchy, i.e., the model hierarchy is strict.

These three predictions for the economically organized dictionary are summarized in Fig. 2c, which shows how much any given level j (on the y-axis) combinatorially contributes to level i (on the x-axis). The first prediction—that there should be about seven levels—is illustrated in the figure by the fact that the matrix is 7 by 7. The second prediction—that the level–level combinatorial growth exponent should be approximately 1.3—is shown in each square of the matrix. Finally, the fact that the predicted hierarchy is strict—where each level contributes to the definitions of words just one level above its own level—is indicated by the contributions to the matrix in Fig. 2c being one below the diagonal.

Next I set out to test these predictions. Section 3 describes the measurements I made from WordNet and the Oxford English Dictionary, and Section 4 sets out to address whether these dictionaries have the above three signature features of an economically organized dictionary.

3. Methods

In order to determine whether actual dictionaries have the signature features of an economically organized dictionary, we need to identify hierarchical levels in the dictionary, and determine the manner in which words of any level are employed in the definitions of words at other levels. In a dictionary hierarchy, words at higher hierarchical levels are more concrete, or more specific, or less fundamental, than the words lower in the hierarchy. For the purpose of determining in what hierarchical level a dictionary word lies, the notion of *hypernym level* was utilized as a measure of how specific a word is. When a B is a kind of C, it is said that C is a *hypernym* of B. For example, 'vehicle' is a hypernym of 'car' and 'train'. The hypernym of a word is less specific, or more generic, or more basic, than the word. Some words have no hypernyms, and are in this sense the least specific or most basic; these words may be said to have a hypernym level of 0. Words having one of these level-0 words as a hypernym have a hypernym level of 1. And, generally, a word's *hypernym level* is the number of steps in this hypernym tree it takes to get from the word to a level-0 word. I use hypernym level as an operational measure of the level of concreteness of words, and as a proxy for hierarchical level.

Used in this study were the hypernym trees created for the English language via the laboratory of George A. Miller, available through WordNet (Fellbaum, 1998). For example, in this tree, level-0 words include 'entity', 'psychological feature', 'abstraction', and 'state' (see Appendix B for all of them). I created my own software that, in combination with the WordNet software and WordNet database files, computed the hypernym level of each of the nouns in WordNet. (Throughout this paper, by 'word' I almost always mean 'noun', which also includes phrases such as "american_bison" or "alley_cat", treated as separate entries in WordNet.) In a small percentage of cases there is more than one hypernym path to a root (meaning the hypernym tree is not, strictly speaking, a tree), and in these cases I assumed that the level of concreteness was represented best by the maximum distance path to a root. In this way, for each of the approximately 141,000 nouns in WordNet I measured its "level of concreteness," or "level of specificness." Hypernym levels in WordNet range from 0 to 17, but because there were only three words in level 17, I confined my analysis to levels 0–16. The distribution of hypernym levels is shown in Fig. 3, along with example words from each level. It is important to emphasize that hypernym level serves only as an operational measure, or a proxy, of the level of concreteness of words. Two words on different branches of the hypernym tree sharing the same hypernym level could nevertheless differ in their concreteness level, for it could be that the dictionary happens to have more finely grained categorical classes along one branch than the other. And the hypernyms in WordNet are by no means unambiguous, depending to some extent on the lexicographer's intuitions. It is reasonable, however, to expect that hypernym level correlates with concreteness level, and this motivates its use here.

With hypernym level as the operational measure of the level of concreteness of words, and hierarchical level, we are, as mentioned earlier, interested in measuring how the hypernym levels of words in a definition relate to the hypernym level of the defined word. One difficulty in carrying out this measurement is that words appearing in definitions typically have multiple senses, and the different senses often differ substantially in their hypernym level. Although the intended sense is almost always unambiguous to a human reader given the context of the definition, the task of determining the intended sense of a word is not easily susceptible to computer automation. Automatic methods are being attempted by the lab of Moldovan (Moldovan & Novischi, 2004) using a set of heuristics; however, such techniques are currently useful only for the words in the glosses of WordNet, not for the words in other definitions, like the OED. Because (a) I wished to have "ground truth" estimates of the relationship between the hypernym level of a word and that of the words in its definition (without the use of any "black box" disambiguation algorithm), and (b) I am interested in directly comparing the measurements in

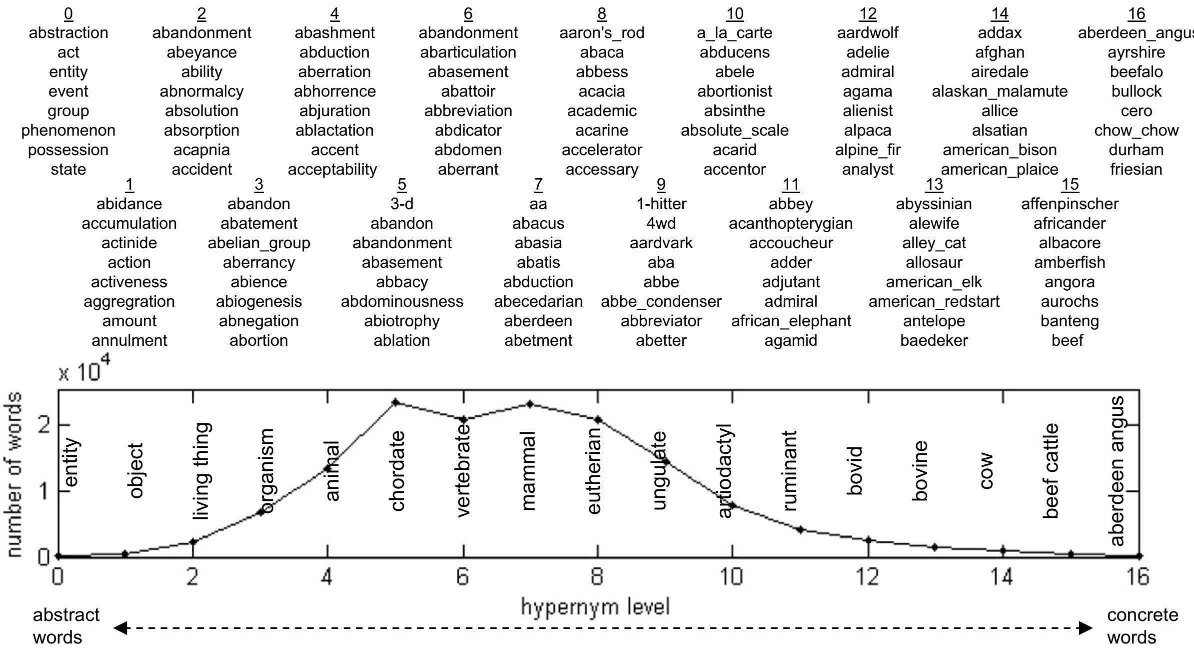

Fig. 3. Illustration of the large range of levels of concreteness for words in English, from abstract words (low hypernym levels, toward the left) to concrete words (high hypernym levels, toward the right). Example words from each hypernym level are shown (the first eight alphabetically from the Oxford English Dictionary, as described in Appendix B). The hypernym level of a word is the number of steps it takes to get from the word, via hypernym connections, to a most-abstract word having no hypernym. The plot shows the number of words for each hypernym level, across 141,755 nouns in WordNet. Within the plot is an example tower of words, showing the successive hypernyms below "aberdeen angus". Note that there are very few words with hypernym levels above about 10, and these words disproportionately concern hoofed animals, dogs and fish, due to farming and domestication; nearly all the words have hypernym levels in the range of 8–10, or lower.

WordNet to that in the OED, I chose to take a *sample* of definitions from WordNet and the OED, and, for each word appearing in the definition, *manually* disambiguate the word's sense.

In particular, for each content word in a definition, or *def-word* (whether in WordNet or the OED), (i) I took the noun form of the def-word if it is not a noun (and sometimes there was no noun form, in which case no data for this def-word was recorded), (ii) I personally determined which of the senses of that def-word in WordNet is the intended one for that definition, a task that is tedious but typically clear (and when at times unclear to me, no data for this def-word was recorded), and (iii) computed that word-sense's hypernym level using the software described earlier. See Fellbaum, Grabowski, and Landes (1998) for evidence that observers can reliably disambiguate word sense. In particular, non-lexicographer observers in the experiments described in that chapter agreed with lexicographers on the appropriate sense of a noun in about 80% of the cases. In addition, when there were only two candidate senses from which to choose, the average agreement was about 85%, whereas when the number of senses increased to eight or more the agreement was still 70–75%, which means that in the latter conditions, naïve subjects are doing around 6 times (or greater) above chance. Furthermore, these success rates of about 80% are lower estimates, because even multiple lexicographers will disagree with one another in some percentage of the cases. For example, if a lexicographer disagrees with another lexicographer 10% of the time, then the naïve 80% success rate needs to be compared to the "ideal" of 90%, not 100%.

Data were collected from two distinct sources of definitions, namely WordNet (which has short definitions called "glosses", and for which I used only the portion of the gloss before the semicolon, after which WordNet typically gives examples of use) and the Oxford English Dictionary, Second Edition (where only the main definition is used, not parenthetic remarks, notes on the plural version or variants, or descriptions of use). The words were sampled by taking, for each hypernym level, the first 30 words occurring alphabetically in WordNet of that hypernym level. That is, I always used WordNet to choose the sample of words, even when the definitions of the words were measured from the OED. Some word entries in WordNet did not exist in the OED (e.g., many entries in WordNet consist of multiple words, such as "blue marlin"), and when this was the case, the alphabetically next word of the appropriate hypernym level was sampled from WordNet, until the definitions from 30 words per level were measured from the OED. For two hypernym levels, the OED data possess fewer words than that of WordNet: First, due to a paucity of level-0 words, only ten words of hypernym level 0 were sampled from WordNet, and only eight of these had unambiguous definitions from the OED. And second, only 20 level-16 words from WordNet could be found in the OED. In total, then, the definitions of 490

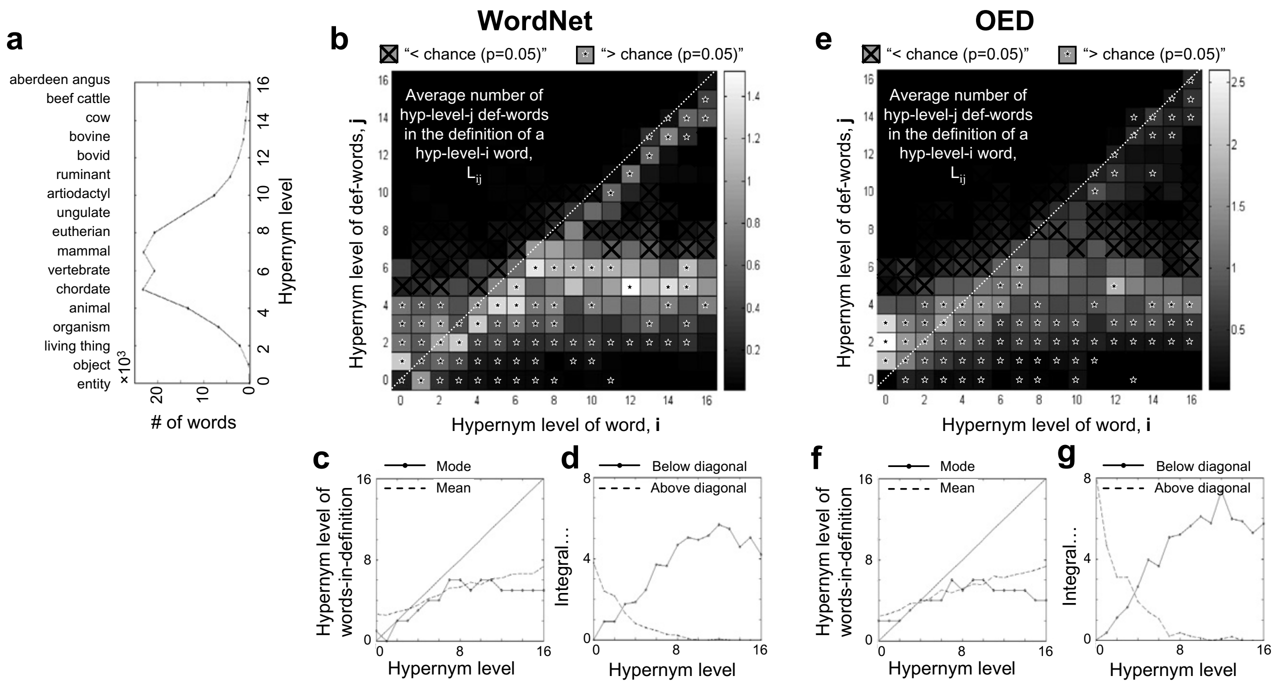

Fig. 4. (a) Number of words in each hypernym level, across all 141,755 nouns in WordNet, as shown in Fig. 3. It is rotated 90° counter-clockwise to help in understanding (b) and (e). Also shown are one example word for each hypernym level, starting from ‘aberdeen angus’. (b, e) Dictionary definition matrix measured from the glosses in WordNet and the definitions in the OED. “*”s (“X”s) are placed on all positive values which are greater (less) than that expected by chance [at the $p = 0.05$ level, assuming the same number of def-words were picked randomly from the distribution in (a)]. One can see from the significance tests that, at each hypernym level i for words, the def-word hypernym-level distributions deviate significantly from the overall distribution in (a). These plots provide an impression of the overall organization of these dictionaries, and in particular their *definition networks*, where a word C has an arrow in the network to word B just in case B occurs in the definition of C. [Note that definition networks are quite different from semantic networks where edges between words are due to relationships such as synonymy, antonymy, hyponymy, hypernymy, meronymy and holonymy (e.g., see Ravasz and Barabasi, 2003; Sigman and Cecchi, 2002; Steyvers and Tenenbaum, 2005).] (c, f) Mode (solid line with data points) and mean (dashed line) hypernym-level of def-words versus the hypernym level of the defined word, for WordNet and the OED. One can see that the mode and mean are by no means flat (WordNet mode $R^2 = 0.60$, df = 15, $t = 4.76$, $p < 0.0005$; WordNet mean $R^2 = 0.96$, df = 15, $t = 20.25$, $p < 10^{-6}$; OED mode $R^2 = 0.40$, df = 15, $t = 3.17$, $p < 0.01$; OED mean $R^2 = 0.98$, df = 15, $t = 26.00$, $p < 10^{-6}$). Each also tends to fall below the diagonal (except at the lower levels). (d, g) The average number of def-words below (solid line) and above (dashed line) that of the defined word (i.e., below and above the diagonal in (b) and (e)). Here we see no range of levels where “below” and “above” maintain approximately equal values; instead, they rapidly cross over at approximately hypernym level 3 (nearer to 2 for WordNet, and approximately 4 for OED). Note that the largest sized hypernym levels are levels 5, 6, 7 and 8 [see (a)], and thus the transition from “above diagonal” to “below diagonal” occurs at a level lower than that of the bulk of the English vocabulary.

words were measured from the glosses in WordNet, and the definitions of 478 words were measured from the OED. See Appendix B for a full list of the words sampled from each. The total number of words in definitions sampled from definitions in WordNet was 2288, and the total for the OED was 3012.

The measured relationship between the hypernym level of words and that of their definition words, or “def-words,” is shown in Fig. 4b and e for WordNet glosses and OED definitions, respectively. Matrix element L_{ij} shows the average number of def-words of level j that occurs for words of level i. Values below the diagonal are, then, cases where the def-word hypernym level is below (i.e., less concrete than, less specific than, or more fundamental than) that of the defined word, consistent with a reductionistic dictionary. A comparison of Fig. 4b and e reveals that they look very similar, despite large differences in the history and methodology underlying how each was built. It suggests that the principles governing the relationship between the level of concreteness of a word and that of its def-words is robust. We make several empirical observations before moving on in the next section to an analysis of whether these dictionaries possess the three signature features of an economically organized dictionary (as discussed in Section 2).

First we ask whether the results in Fig. 4b and e are a consequence of some fairly simple null hypothesis? I examined three null hypotheses. The most obvious null hypothesis to consider is that the hypernym levels of words found in definitions are sampled randomly from the overall distribution of hypernym levels shown in Fig. 4a (and also shown in Fig. 3). If this were the case, then each vertical column in Fig. 4b and e would be statistically indistinguishable from the distribution in Fig. 4a. Instead, for each hypernym level, i, of a word (i.e., for any column in Fig. 4b and e), the distribution of levels of the def-words significantly deviates from the distribution in Fig. 4a. A second natural null hypothesis to test is that the distribution of hypernym levels for def-words tends to stay constant, no matter the level of the defined word. If this were the case, then each of the columns in Fig. 4b and e would look the same. This is not the case. Instead, the mean and the mode hypernym level of def-words increase as the hypernym level of the defined word increases (i.e., as i increases), as shown in Fig. 4c and f. A third potential null hypothesis is that the average hypernym level of def-words tends to be near that of the defined word, sometimes more concrete (more specific), sometimes less concrete (less specific), but on average the same level of concreteness. If this were the case, then the matrices in Fig. 4b and e would have contributions only on or near the diagonal, and tend to be above the diagonal (def-word is more concrete than the defined word) as often as below the diagonal (def-word is less concrete than the defined word). Examination of the matrices reveals that this is not the case, and Fig. 4d and g shows that there is no extended range of hypernym levels for defined words where the weight above the diagonal approximately matches that below the diagonal.

None of these three null hypotheses, then, explains the organization of the dictionary shown in Fig. 4b and e. Instead, one of the more salient features is that over most of the range of hypernym levels of defined words (i.e., for most values of i), the hypernym levels of the def-words tend to be below that of the defined word. That is, most of the contributions in the matrices of Fig. 4b and e are below the diagonal, and many of these contributions are *significantly* more than expected by chance if pulling from the overall distribution in Fig. 4a (as shown by a '$*$'). Furthermore, many of the contributions above the diagonal are significantly lower than expected by chance (as shown by an 'X'). This approximately describes these matrices for hypernym levels of defined words from about level 4 and up. In fact, Fig. 4e and 4g indicate that after about level 3 or 4, the def-word hypernym levels below the diagonal outweigh those above the diagonal. More than 90% of the words in the WordNet dictionary have a hypernym level of 4 or higher (see Fig. 3), and therefore most words have def-words that are less specific (or more basic) than themselves, consistent with what a reductionistic model would expect. This is not empirically surprising because dictionary definitions typically refer to a *genus*, which is the hypernym of the word, and our measurements tended to concentrate on the genus and avoided the differentia (e.g., by taking only the first portion of WordNet definitions, and not including examples in the OED).

However, there are non-reductionistic features evident in Fig. 4b and e.

The first non-reductionistic feature is that at all hypernym levels (i.e., all columns) there are typically a small number of def-words coming from on or above the diagonal. There are two main reasons why our data would be expected to have this feature even if the dictionary did not. (a) Dictionary definitions sometimes give examples of the word (even though the data collection tended to avoid this), which are therefore more specific, and will have a tendency to lie above the diagonal. Furthermore, dictionaries often note related words of the same level of concreteness, which will have a tendency to lie along the diagonal. If this point is true, then we would expect this point to apply to the Oxford English Dictionary to a greater extent than to WordNet glosses, because the latter does not tend to include examples or note related words (especially the portion of the gloss before the semicolon, which tends to focus on a terse definition). This is indeed the case as can be seen by comparing Fig. 4b and e. (b) Also, one must recall that hypernym levels are only an operational measure of the level of concreteness of words, and can only be expected to correlate with the true level of concreteness. Therefore, even if for some level the definitions are purely reductionistic (i.e., below the diagonal in the matrix) in the dictionary, our measurements from the dictionary may have some non-reductionistic contributions (i.e., on or above the diagonal). This is especially true if there are significant contributions from the level one below that of the defined word—as is the case in our data—for then discrepancies between our hypernym-level proxy and the true level of concreteness have a greater chance of spilling onto or above the diagonal.

The second non-reductionistic feature found in the dictionary matrices of Fig. 4b and e is that there are more non-reductionistic (on or above the diagonal) contributions at the lower hypernym levels than at the other levels. (See Calzolari, 1988, for early hints that there is a qualitative change in the definitions amongst the most basic, or least specific, words.) Our measurements would be expected to have this feature—even if the dictionaries were always purely reductionistic (i.e., below the diagonal)—because dictionaries do not have word entries without giving *some* discussion of the word's semantics and use. That is, unlike the toy dictionaries in Fig. 1 where "0" and "1" are left undefined, real dictionaries will not leave them undefined. If it is not possible to give the meaning of a lower-level word in terms of other words, the dictionary would *still* give examples of the word, describe how it is used, and note its relations to other words, which would lead to words in the dictionary definition that have a tendency to be on or above the diagonal. Again, if this is true, then we would expect this to apply to the definitions in the Oxford English Dictionary to a greater extent than to the glosses in WordNet, and indeed this is the case. For example, Fig. 4c and f shows that the mode level for def-words is on or above the diagonal only for level 2 in WordNet, whereas this occurs for levels 1 through 4 in the OED (not counting level 0, which by necessity must have a contribution on or above the diagonal). Also, Fig. 4d and g shows that the point at which below-diagonal contributions begin to outweigh above-diagonal contributions occurs between levels 2 and 3 for WordNet, but at nearly level 4 for the OED. There is, furthermore, another reason to expect these non-reductionistic features at these low levels, and it is that the choice of which words are to be the atoms has arbitrariness (as evidenced, for example, by the range of 10–60 atoms), and given the small number of words at these lower levels, signs of reductionistic definitions are lost.

Therefore, despite our measurements having some features that are not reductionistic, they occur just where one would expect even if dictionaries were reductionistic. If, on the other hand, dictionaries were severely non-reductionistic in their organization, then one might expect to find that our measurements would have been more thoroughly non-reductionistic, rather than in just the two ways discussed above. Together, this is strongly suggestive that dictionaries are reductionistic. On the one hand this should not come as an empirical surprise, as mentioned earlier; however, there is a common misconception that dictionaries are deeply circular, something these data argue against.

A final salient feature found in the measurements shown in Fig. 4b and e is that at around level 9 and above, the def-word distributions become bimodal, with a lower distribution remaining approximately constant, and the upper distribution being centered approximately one level below the diagonal. That there are significant contributions from the latter is not surprising, because definitions will often refer to a word's hypernym. What is not necessarily expected is (i) that the lower distribution remains approximately constant, and centered at approximately level $j = 5$, and (ii) that there are not many contributions from hyponyms (if B is a hypernym of C, then C is a hyponym of B) which would be above the diagonal. This regime of the plot may be an artifact of animal domestication, for most of the highest hypernym level words are cows, horses, goats, dogs and fish, where the English language has acquired a hyper-refined hierarchical scale, like the one shown for 'Aberdeen Angus' in Fig. 3. At any rate, this regime probably is of little importance compared to the below-level-9 regime for several reasons. First, note that there are relatively few words here, namely only 12% of all 141,000 in WordNet (see Fig. 3). Second, unlike the low-hypernym-level words, for which there are also few words, these high-hypernym-level words play little or no role in the definitions of other words. This can be seen by noting the absence of matrix values in the upper middle and upper left of Fig. 4b and e. Finally, the fact that they are defined mostly by words at a constant level, namely about $j = 5$, suggests that despite ranging in hypernym level from 9 through 16, they may fundamentally be at a similar level of concreteness.

4. The dictionary has the signature of an economically organized hierarchy

In order to adequately assess the extent to which the actual dictionary's organization possesses the three signature features of the model economically-organized dictionary, we must measure how the dictionary combinatorially grows from level to level. The organization of the actual dictionary is not a clean, strict hierarchy with a single level–level combinatorial growth exponent applying between every pair of adjacent levels like in the model (summarized in Fig. 2c). Instead, for each pair of hypernym levels, i and j, it is necessary to measure the extent to which level j combinatorially contributes to defining the words in level i. Rather than a single level–level combinatorial growth exponent, d, as discussed in Section 2 for the model, there is a matrix of d_{ij} values, where d_{ij} is the level–level combinatorial growth that level j contributes to the construction of level i. See Appendix A for discussion of how these d_{ij} are computed.

Fig. 5a and d shows the level–level combinatorial growth matrices for WordNet and the OED. As was the case for Fig. 4b and e, the matrices look very similar. The question is, Do these level–level combinatorial growth matrices have the three signature features of an economically organized dictionary, as summarized in Fig. 2c? (Note that it is doubtful that these dictionaries are actually globally optimal, because (i) there are other selection pressures shaping their organization and (ii) the mechanisms shaping the organization of the dictionary are unable to find the global optimum.)

4.1. First signature feature

The first signature feature was that the predicted combinatorial hierarchy possesses approximately in the range of 5–7 levels (given the plausible range of 10–60 for the number of bottom-level words), although hierarchies with about 4–10 levels are within 10% of optimal. From an initial glance at Fig. 5a and d one might conclude that there are 17 levels (0 through 16), which is well above the range expected for an economically organized dictionary. However, as discussed in Section 3, levels above 9 may be of little significance in understanding the general principles of the dictionary organization. In fact, as we will see in a moment, although the definitions tend to be reductionistic above level 9, words in these levels are not participating in the combinatorial hierarchy of the dictionary.

Which levels, if any, *are* participating in the combinatorial hierarchy, where again the economical dictionary would predict that there are about 5–7, and that there are from 4 to 10 if near-optimal? To begin to answer this, we must distinguish between two distinct kinds of combinatorial growth information in which one may be interested. The first concerns how combinatorially the level was *built* from other levels. It is called the *receiving-combinatorial-growth exponent*, written as d_i, and is the sum of the d_{ij} elements in *column* j of the matrix. The second kind of combinatorial growth information about a level concerns how combinatorially that level is *used* to build other levels. It is called the *contributing-combinatorial-growth exponent*, written as d_j, and is the sum of the d_{ij} elements in a *row* of the matrix. In the model discussed earlier, these two distinct concepts happened to coincide because the model assumed that hierarchy was strict and the level–level combinatorial growth exponent is always the same between any pair of adjacent levels. For the data, however, this is not the case, and these two quantities must be distinguished.

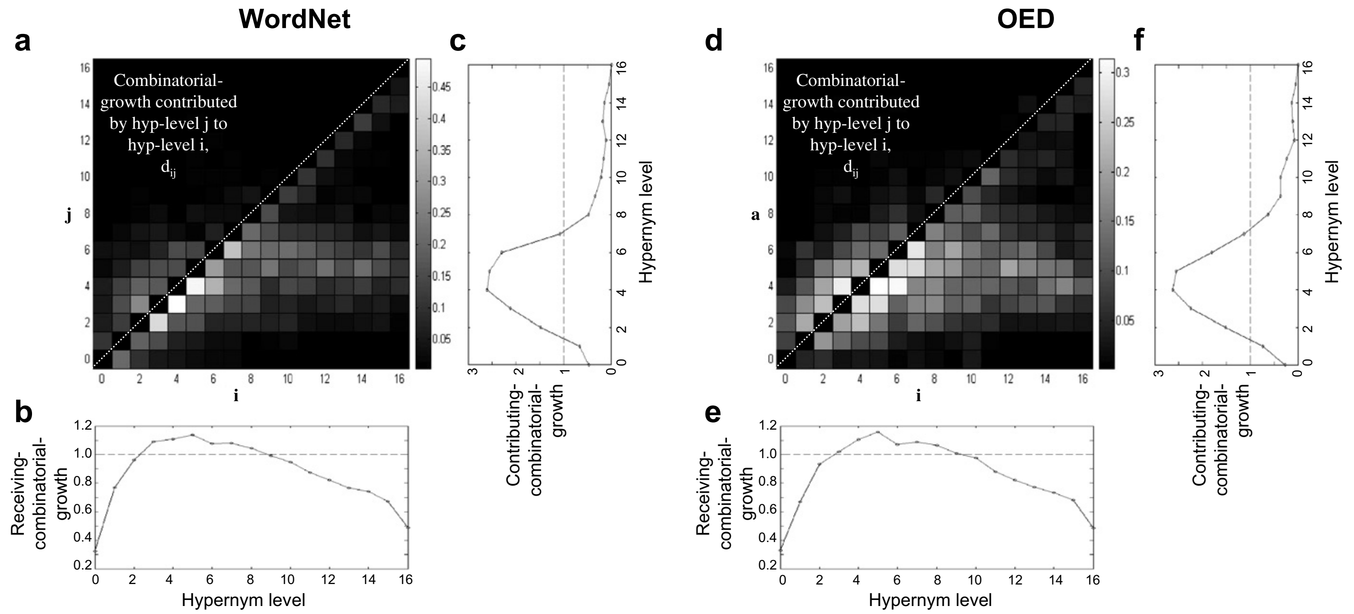

Fig. 5. Data testing the prediction in Fig. 2c. (a, d) The combinatorial-growth matrix, d_{ij}, for WordNet and the OED. Intuitively, d_{ij} is the degree to which hypernym-level j is combinatorially employed in the construction of hypernym level i. (See Appendix A for its definition.) One can see that the large combinatorial-growth elements occur approximately in the range from levels 3–7, and are clustered just below the diagonal. These two empirical plots were theoretically predicted in Fig. 2. (b, e) The receiving-combinatorial-growth, d_i, for each hypernym level, which is, for each level i, the sum of the d_{ij} in column i. d_i indicates how combinatorially hypernym level i is built (from all other levels). $d_i > 1$ implies that level i is built combinatorially out of the words in the levels that contribute to it; $d_i \leqslant 1$ implies that level i is not built combinatorially. One can see that in both WordNet and the OED, only hypernym levels 3 through 8 are combinatorially built, having a broad plateau with a soft peak at level 5 in each case. One can also see that the receiving-combinatorial-growth values are not much above one, meaning that these combinatorially-built levels are not very combinatorial at all. (c, f) The contributing-combinatorial-growth, d_j, for each hypernym level, which is, for each level j, the sum of the d_{ij} in row j. d_j indicates how combinatorially hypernym level j *contributes*to all other levels. $d_j > 1$ implies that the words in hypernym level j are combinatorially harnessed to build other levels; $d_j \leqslant 1$ implies that level j is not so combinatorially harnessed. One can see that the contributing-combinatorial-growth values are more sharply peaked than the receiving-combinatorial-growth values, in each case with a peak value at level 4, and $d_j > 1$ for levels 2 through 7. Note that this set of contributing levels is the same as the set of receiving levels, but decremented by one.

First let us ask which levels are *built* combinatorially. A level is built combinatorially if the receiving-combinatorial-growth exponent, d_i, is greater than one. And the larger the exponent, the more combinatorially it is built out of other (typically lower) levels. The receiving-combinatorial-growth exponents for each level are shown in Fig. 5b and e. As one can see, the receiving-combinatorial-growth exponents are greater than one for levels 3 through 8, meaning that there are six adjacent levels in the hierarchy that are built combinatorially.

Now let us ask which levels are *used* combinatorially. A level is used combinatorially (to build other, typically higher, levels) if the contributing-combinatorial-growth exponent, d_j, is greater than one. The contributing-combinatorial-growth exponents for each level are shown in Fig. 5c and f. They are greater than one for levels 2 through 7, and so there are six adjacent levels that are combinatorially used to build other levels.

Therefore, for both WordNet and the OED, the six levels that are *used* combinatorially are levels 2 through 7, whereas the six levels that are *built* combinatorially are levels 3 through 8. Importantly, this means there is a combinatorial hierarchy from level 2 through 8, making seven levels in all. This fits well within the predicted range of levels for an optimally organized dictionary, which was from 5 to 7, and from 4 to 10 for those within 10% of optimal. The seven levels in this combinatorial hierarchy (from 2 through 8) account for 110,000 words of the 141,000 in WordNet, or 78%. For example, natural kind terms, or "basic terms" (Rosch, 1978)—such as 'car' (level 10), 'chair' (level 8), 'table' (level 7), and 'lamp' (level 7)—tend to be approximately at the top of this hierarchy, whereas superordinate terms (e.g., 'furniture' at level 6) are at lower levels.

What about the levels outside of this range? The upper, more specific, levels 9 through 16 do not participate in the combinatorial hierarchy (because their receiving- and contributing-combinatorial-growth exponents are below one), but this is what we already expected, as discussed earlier. What about the lowest levels, namely 0 and 1, which also do not appear to be part of the combinatorial hierarchy? First, we must remember from Fig. 3 that there are only a relatively small number of words in these two levels, namely 281 words, or 0.2% of all 141,000 words in WordNet. Second, as discussed in Section 3, even if the dictionary is economically organized all the way to the bottom, we would expect our measurements to deviate from this as seen here. And this is true even assuming that hypernym level is a perfectly accurate measure of the level of concreteness, something that is almost certainly not the case. If levels 0 and 1 actually are part of the combinatorial hierarchy in the dictionary, then there would be nine levels (i.e., levels 0 through 8), still within the range of near-optimal number of levels (which was from 4 through 10).

Before moving to test whether these dictionaries possess the other two signature features of an economically organized dictionary, it is revealing to look into estimates of the number of hierarchical levels for the lexicon from lexicographic researchers within the natural semantic metalanguage community who have over forty or more years studied semantic primitives and how they combine to adequately give meanings to all the words in the lexicon (Goddard, 2006; Goddard & Wierzbicka, 2002; Wierzbicka, 1996; see also related work from a different school in, e.g., Apresjan, 2000, especially chapter 8). Rather than estimating the number of hierarchical levels by quantitatively analyzing the global organization of dictionaries as I do here, these researchers have used their vast knowledge of the lexicon and lexical relationships across many languages to make plausible estimates of the number of hierarchical levels.

For example, Goddard (2007) estimates that there are

> "as many as four levels of semantic nesting within highly complex concepts, such as those for natural kinds and artefacts. In the explication for cats or chairs, for example, the most complex molecules are bodily action verbs like 'eat' or 'sit'. They contain body-part molecules such as 'mouth' and 'legs'. These in turn contain shape descriptors, such as 'long', 'round, and 'flat', and they in turn harbour the molecule 'hands', composed purely of semantic primes." (Goddard, 2007, p. 10)

Recall from Section 2 that, for the natural semantic metalanguage (NSM) approach to semantics, *semantic primes* are the atomic (or bottom-level) words in a hierarchy. Also, *semantic molecules* are words built from semantic primes that are, in turn, used to define higher-level words (Goddard, 2007). [Semantic molecules as defined by Goddard (2007), are under a further constraint above and beyond what is required here for a word to be at an interemediate level in the hierarchy. Something is a semantic molecule only if "it emerges from the analytical process that the required semantic content cannot be represented directly in an intelligible fashion using semantic primes" (Goddard, 2007). Intermediate-level words as I treat them may or may not satisfy this.] Goddard's judgment of "four levels of semantic nesting within highly complex concepts" means that, with the addition of the level of the highly complex concept itself, he concludes that there are five levels in our terms.

Wierzbicka comes to a similar conclusion in a section titled "The hierarchical structure of the lexicon" (Wierzbicka, in press).

> "The molecular structure of the lexicon has not yet been investigated for a long time and much remains to be discovered. It is already known, however, that there are several levels of molecules: those of level one (M_1) are built directly of semantic primes, those of level two (M_2) include in their meaning molecules of level one, [and so on]. It is very likely that there are also molecules of level four and five. ...
>
> A more complex sequence is built on the concept 'niebo' ('sky'). 'Niebo' itself, which is an M_1, generates, as it were, molecules like 'stońce' ('sun'), 'gwiazda' ('star'), 'księżyc' ('moon'), 'chmura' (roughly, 'dark cloud'), 'zorza' ('aurora') and 'neibieski' (roughly, 'blue')—each an M_2. 'Stońce', in turn, generates 'dzień' ('day'), which is an M_3, and which is included, for example, in the meanins of words like 'poniedziatek' ('Monday'), 'wtorek' ('Tuesday'), and so on." (Wierzbicka, in press, p. 7)

Her approximate estimate of molecules of level four or five translates (once we include the semantic primes as the final level) into 5 or 6 levels, similar to the estimate above by Goddard.

These two estimates—of about 5 or 6 levels—based on the experiences of seasoned lexicographers is broadly consistent with the order of magnitude of the number of hierarchical levels I have empirically found here—namely about 7—above for WordNet and the OED. And, these estimates are also consistent with the prediction from my model of about 5–7 hierarchical levels (from Section 2 and Fig. 2c).

4.2. Second signature feature

The second signature feature of an optimally organized dictionary was that the level–level combinatorial growth exponent is low, namely in the approximate range from 1.2 to 1.5. Recall that for the model, the receiving-combinatorial-growth and the contributing-combinatorial-growth exponent were conflated into a single number, namely given by the level–level combinatorial growth exponent. So, to answer whether the actual dictionary's combinatorial growth exponents fall into this predicted range, we need to look at both the receiving- and contributing-combinatorial-growth exponents. Among the receiving-combinatorial-growth exponents greater than one (i.e., confining analysis to the combinatorial hierarchy from levels 2 through 8), the averages are 1.09 for WordNet and also 1.09 for the OED. Similarly, for the contributing-combinatorial-growth exponents the averages are 2.02 and 1.96, respectively. (I note as an aside how, unlike the receiving-combinatorial-growth exponents which are relatively flat and low over the range of values greater than one, the contributing-combinatorial-growth exponents vary considerably over the range, suggesting that words at some levels—namely levels 3 through 5—are better suited at defining other words.) For the purposes of comparing a single number to the predicted level–level combinatorial growth range of 1.2–1.5, I took the average of the average receiving- and contributing-combinatorial-growth exponents, and accordingly get 1.54. This is very close to the predicted range of level–level combinatorial growth exponents for the optimal dictionary.

4.3. Third signature feature

The third and final signature feature for an economically-organized dictionary was that the dictionary's definitions be arranged in a strict hierarchy, where each level contributes only to the definitions of the level directly above it. In terms of the combinatorial growth matrices in Fig. 5, this would correspond to all the contributions coming from the elements one below the diagonal as shown in Fig. 2c. One can see that this is approximately the case in Fig. 5a and b, at least from about hypernym level 2 to level 7, where the largest value (or whitest) tends to be one below the diagonal. This is especially true for WordNet, which is what we would expect because, as discussed earlier, it is more succinct, refers to fewer or no examples, and does not inform us about use.

5. Conclusion

To sum up, then, WordNet and the Oxford English Dictionary possess all three signatures of an economically organized dictionary. First, they possess a combinatorial hierarchy with around seven levels, which fits within the predicted range of 5–7 for an optimal hierarchy. (And if levels 0 and 1 are part of the hierarchy in the dictionary despite not measuring as so in my measurements, then there are nine levels, still within the near-optimal predicted range of 4–10.) Second, their combinatorial growth exponents are around 1.54, close to the predicted range of 1.2–1.5. And third, the combinatorial hierarchy has a tendency to be strict—i.e., words tend to be used to define words in the level just above them—as in the model economical dictionary. In short, the signature features of an optimal hierarchy as shown in Fig. 2c were found to be present in these dictionaries, as seen in Fig. 5a and 5d, providing support for the hypothesis that dictionaries like WordNet and the Oxford English Dictionary are organized in such a way as to economize the amount of dictionary space required.

One speculative hypothesis is that the signature of an economically organized hierarchy we find in these dictionaries is not for a smaller, more economical dictionary, per se, but because the lexicon itself has evolved over tens of thousands of years by cultural selection to help lower the overall "brain space" required to encode the lexicon. Many of our other human inventions have been designed—either explicitly or via cultural selection over time—so as to minimize their demands on the brain. For example, writing and other human visual signs appear to have been optimized by cultural selection for our visual systems (Changizi, 2006, 2009; Changizi & Shimojo, 2005; Changizi, Zhang, Ye, & Shimojo, 2006). The definitions in the dictionary are not identical to the meanings of words we have in our heads, missing out, for example, on metaphorical associations that may be part of an individual's meaning of the word (see, e.g., Fillmore, 1975; Lakoff & Johnson, 1980, 2003), but it would be surprising if the large-scale organization of the dictionary was not driven in some large part by the organization of our mental lexicon.

Acknowledgement

Support for this research was given by NIH Grant 5F32EY015370-02.

Appendix A. Computing how hypernym level *j* combinatorially contributes to defining level *i*

Fig. 4b and e shows how the hypernym level of a word relates to the hypernym levels of its def-words, as described in the main text. Intuitively, it amounts to a connectivity matrix, where the hypernym levels are the nodes. However, because this does not take into account the sizes of the levels, this information does not capture the "combinatorial contributions" levels make to the construction of other levels, and it is this kind of combinatorial growth information that is crucial to testing the predictions summarized by Fig. 2c.

Measuring the level–level combinatorial growth for the model discussed in Section 2 (and summarized in Fig. 2c) is simple because a single number, d, characterizes it, and this is due to the hierarchy being strict (i.e., each level contributes to the level just above it), and due to the growth from D_i to D_{i+1} being the same for all i. Characterizing the level–level combinatorial growth for the actual dictionary is more complicated because the hierarchies are not entirely strict (i.e., multiple levels contribute to the definitions of another level), and the growth from one level to the next is not always the same across i.

If the dictionary *were* strict like the model, then as discussed in Section 2, the relationship between, say, levels 0 and 1 would be given by

$$D_1 = D_0^{\mathrm{d}}. \tag{1}$$

The level–level combinatorial growth exponent is simply

$$d = (\log D_1)/(\log D_0). \tag{2}$$

But now let us suppose that levels 0 and 1 are used to build level 2, and that the sizes of the levels are D_0, D_1 and D_2, respectively. The appropriate generalization of Eq. (1) is,

$$D_2 = (D_0^{d_{2,0}})(D_1^{d_{2,1}}), \tag{3}$$

where d_{2j} is the combinatorial exponent quantifying the extent to which level j combinatorially contributes to the definitions for words in level 2. Taking the logarithm, we have

$$\log(D_2) = d_{2,0}\log(D_0) + d_{2,1}\log(D_1).$$

Let L_2 be the average number of words in a definition of a level-2 word; $L_{2,0}$ of the words per definition are from level 0, and $L_{2,1}$ are from level 1. These "L" data are what are shown in Fig. 4b and e. Because of redundancies, the number of degrees of freedom in using D_0 and D_1 to build D_2 may be lower than L_2, and let β_2 be that fraction, so that $d_{2,j} = \beta_2 L_{2,j}$. Eq. (3) can now be manipulated into

$$\log(D_2) = \beta_2[L_{2,0}\log(D_0) + L_{2,1}\log(D_1)]$$

and thus

$$\beta_2 = \log(D_2)/[L_{2,0}\log(D_0) + L_{2,1}\log(D_1)].$$

Because $d_{2,j} = \beta_2 \mathrm{L}_{2,j}$, with a little algebra we conclude that

$$d_{2,\mathrm{j}} = \log(D_2)/[(L_{2,0}/L_{2,\mathrm{j}})\log(D_0) + (L_{2,1}/L_{2,\mathrm{j}})\log(D_1)], \quad (4)$$

where j can be 0 or 1 here. This is similar to Eq. (2), except that in the denominator is the sum of the logarithms of the sizes of the contributing levels, relatively weighted by how many words per definition they contribute.

The previous equation is for the case where two levels contribute to define a third level, but from it the fully general equation is easy to see, and is given by

$$d_{ij} = \log(D_i)/\Sigma_{k\neq i}[(L_{ik}/L_{ij})\log(D_k)], \quad \text{where } i \neq j. \quad (5)$$

Each d_{ij} is the combinatorial growth exponent for the contribution from level j to the construction of level i, and these values are shown in Fig. 5a and d.

Appendix B. List of words from which definitions were sampled

B.1. WordNet

Words in the definitions of the following words were measured from the glosses in **WordNet**, followed in parentheses by the sense # in WordNet.

Hypernym level 0: abstraction (6), act (2), entity (1), event (1), group (1), human_action (1), phenomenon (1), possession (2), psychological_feature (1), state (4). **Hypernym level 1:** abidance (2), accumulation (2), actinide (1), action (1), activeness (1), agency (4), aggregration (1), amount (3), annulment (1), antagonism (1), anticipation (4), assessment (4), association (8), assumption (7), attribute (2), being (1), biological_group (1), biotic_community (1), causal_agency (1), chance (4), circuit (6), citizenry (1), cleavage (1), cognition (1), collection (1), community (8), condition (1), condition (2), conflict (4), consequence (1). **Hypernym level 2:** abandonment (1), abeyance (1), ability (3), abnormalcy (1), absolution (1), absorption (6), acapnia (1), acathexia (1), accenting (1), accident (2), accompaniment (1), account (1), achievement (1), acme (1), acquaintance (2), action (5), actuality (1), actus_reus (1), addiction (1), address (3), adeptness (1), adjudication (1), adroitness (1), adulthood (2), aestivation (2), affair (3), affect (1), affiliation (1), affinity (5). **Hypernym level 3:** abandon (2), abatement (1), abdication (1), abelian_group (1), aberrancy (1), abidance (1), abience (1), ability (1), abiogenesis (1), abnegation (1), abocclusion (1), abortion (2), aboulia (1), about-face (2), abrachia (1), abstractedness (1), abuse (3), abutment (1), ac (1), acardia (1), accelerator (3), acceptance (1), acceptor (1), accession (1), accession (4), accommodation (5), accord (2), achylia (1). **Hypernym level 4:** abashment (1), abdomen (1), abduction (2), aberration (2), abhorrence (1), abjuration (1), ablactation (1), accent (2), acceptability (1), accession (2), accession (6), accessory (2), acclimation (1), accommodation (6), accordance (2), accouchement (1), accretion (2), acetal (1), acetic_anhydride (1), achondrite (1), achromatism (1), acid_anhydrides (1), acnidosporidia (1), acoustic_projection (1), acquired_taste (1), acrasiomycetes (1), acroanaesthesia (1), action (3), action (9), acuteness (1). **Hypernym level 5:** 1-dodecanol (1), 3-d (2), 365_days (1), aar (1), aba (2), abandon (1), abandonment (3), abasement (2), abb (1), abbacy (1), abdominousness (1), abies (1), abiotrophy (1), ablation (1), ablation (2), ablepharia (1), abo_blood_group_system (1), abode (1), abomination (1), abortion (1), about-face (1), abramis (1), abrasion (2), abrasiveness (2), abscess (1), abseil (1), absolute_frequency (1), absolute_zero (1), absolutism (5), absolutism (6). **Hypernym level 6:** 1st-class_mail (1), abamp (1), abandonment (2), abarticulation (1), abasement (1), abattoir (1), abbreviation (1), abcoulomb (1), abdicator (1), abdomen (2), abdominoplasty (1), abecedarian (2), abelmoschus (1), aberrant (1), abfarad (1), abhenry (1), abhorrer (1), abila (1), abkhaz (1), ablative (1), ablaut (1), abode (2), abominator (1), aborigine (1), abrasion (3), abscissa (1), ache (1), achromasia (1), acid_test (1), acidophil (1). **Hypernym level 7:** 12-tone_music (1), a-horizon (1), aa (1), aachen (1), aalborg (1), aalost (1), aalto (1), aarhus (1), aaron_burr (1), abaca (1), abacus (1), abadan (1), abasia (1), abatis (1), abduction (1), abecedarian (1), abel (1), abel_janszoon_tasman (1), abelmoschus_esculentus (1), aberdare (1), aberdeen (1), abetalipoproteinemia (1), abetment (1), abidjan (1), abilene (1), abkhaz (1), abolitionist (1), abomasum (1), aborigine (2). **Hypernym level 8:** aalii (1), aaron (1), aaron's_rod (1), ab (4), abaca (2), abbess (1), acacia (1), academic (1), acarine (1), accelerator (4), accelerator_factor (1), accessary (1), abalone (1), abbott_lawrence_lowell (1), abc (1), abductor (1), abductor (2), abecedarius (1), abelard (1), abelia (1), abnegator (1), abortus (1), abseiler (1), absinthe_oil (1), absolute_pitch (1), abstraction (3), abu_dhabi (1), abuja (1), acalypha_virginica (1), acanthocereus_pentagonus (1). **Hypernym level 9:** 1-hitter (1), 4wd (2), aardvark (1), aaron (1), aaron_copland (1), aba (1), abaya (1), abbe (1), abbe_condenser (1), abbreviator (1), abelmoschus_moschatus (1), abetter (1), ablative_absolute (1), abnaki (1), abney_level (1), abominable_snowman (1), abraham (1), absconder (1), absolute_ceiling (1), abstemiousness (1), abutilon_theophrasti (1), aby_moritz_warburg (1), acacia_auriculiformis (1), acacia_farnesiana (1), academic_costume (1), accelerator (1), accent (5), acciaccatura (1), accord (3). **Hypernym level 10:** 4wd (1), a-bomb (1), a_la_carte (1), abandoned_ship (1), abdominal_aorta (1), abducens (1), abele (1), abortionist (1), abrocome (1), abronia_elliptica (1), absinthe (1), absolute_scale (1), acarid (1), accentor (1), accessory_nerve (1), accidence (1), accident_surgery (1), accipiter_cooperii (1), accordian_door (1), acer_argutum (1), acer_campestre (1), acheta_assimilis (1), achras_zapota (1), acoustic_nerve (1), acquisition_agreement (1), acridid (1), acrocomia_aculeata (1), action (7), active_matrix_screen (1), actual_damages (1). **Hypernym level 11:** abbey (2), abies_bracteata (1), acanthisitta_choris (1), acanthophis_antarcticus (1), acanthopterygian (1), acanthoscelides_obtectus (1), accoucheur (1), acer_negundo_californicum (1), acherontia_atropos (1), acris_gryllus (1), acroclinium_roseum (1), actias_luna (1), actinomeris_alternifolia (1), action_officer (1), active_application (1), acute_lymphoblastic_leukemia (1), adder (3), adelges_abietis (1), adjutant (1), admiral (2), aegyptopithecus (1), aelfred (1), aeromedicine (1), african_crocodile (1), african_elephant (1), african_millet (1), agamid (1), agathis_lanceolata (1), ahab (1), ai (1). **Hypernym level 12:** aardwolf (1), abies_alba (1), acinonyx_jubatus (1), acridotheres_tristis (1), acrocephalus_schoenobaenus (1), actitis_hypoleucos (1), adelie (1), admiral (1), adrian (1), aetiologist (1), african_hunting_dog (1), agama (1), agelaius_phoeniceus (1), agkistrodon_contortrix (1), airbus (1), aix_galericulata (1), albert_edward (1), albula_vulpes (1), alectoris_ruffa (1), alienist (1), allegheny_plum (1), alligator_lizard (1), allmouth (1), alopex_lagopus (1), alopius_vulpinus (1), alpaca (3), alpine_fir (1), amarelle (1), amblyrhynchus_cristatus (1), american_antelope (1). **Hypernym level 13:** 1st_viscount_montgomery_of_alamein (1), a._e._burnside (1), abyssinian (1), acipenser_huso (1), aetobatus_narinari (1), african_green_monkey (1), agricola (1), albrecht_eusebius_wenzel_von_wallenstein (1), alces_alces (1), alcibiades (1), alewife (2), alley_cat (1), allosaur (1), alosa_pseudoharengus (1), ambrose_everett_burnside (1), american_elk (1), american_redstart (1), anas_americana (1), anatotitan (1), andrew_jackson (1), angelfish (2), angora (3), anguilla_sucklandii (1), antelope (1), anthony (1), anthony_wayne (1), antigonus (1), antonio_lopez_de_santa_ana (1), antonius (1). **Hypernym level 14:** abudefduf_saxatilis (1), addax (1), adenota_vardoni (1), aepyceros_melampus (1), afghan (5), agonus_cataphractus (1), airedale (1), alaskan_malamute (1), allice (1), alligatorfish (1), alosa_chrysocloris (1), alsatian (1), ambloplites_rupestris (1), ameiurus_melas (1), american_bison (1), american_merganser (1), american_plaice (1), ammotragus_lervia (1), angelfish (1), anoa (1), antilope_cervicapra (1), apogon_maculatus (1), appaloosa (1), arab (2), arctic_char (1), argal (1), armed_bullhead (1), atlantic_halibut (1), attack_dog (1), aurochs (1). **Hypernym level 15:** abramis_brama (1), acanthocybium_solandri (1), affenpinscher (1), affirmed (1), africander (1), albacore (1), alectis_ciliaris (1), amberfish (1), american_flagfish (1), american_foxhound (1), american_pit_bull_terrier (1), amphiprion_percula (1), angora (1), antidorcas_euchore (1), armored_sea_robin (1), atlantic_bottlenose_dolphin (1), aurochs (1), bairdiella_chrysoura (1), banteng (1), barred_pickerel (1), beef (1), bellwether (2), bezoar_goat (1), bighorn (1), black-and-tan_coonhound (1), black_bass (1), blenheim_spaniel (1), bennius_pholis (1), bluefin (2), bluegill (1). **Hypernym level 16:** aberdeen_angus (1), american_water_spaniel (1), ayrshire (1), beefalo (1), bibos_frontalis (1), black_buffalo (1), black_marlin (1), blue_marlin (1), bucking_bronco (1), bullock (1), cavalla (1), cero (1), chow_chow (1), coney (1), creole-fish (1), durham (1), friesian (1), galloway (1), gaur (1), gayal (1), heifer (1), hereford (1), hind (1), jewfish (2), king_mackerel (1), kingfish (2), makaira_albida (1), santa_gertrudis (1), scomberomorus_maculatus (1), springer (2).

B.2. Oxford English Dictionary

Words in the definitions of the following words were measured from the definitions in the **Oxford English Dictionary** (second edition), followed in parentheses by the sense # as listed in WordNet.

Hypernym level 0: abstraction (6), act (2), entity (1), event (1), group (1), phenomenon (1), possession (2), state (4). **Hypernym level 1:** abidance (2), accumulation (2), actinide (1), action (1), activeness (1), aggregration (1), amount (3), annulment (1), assessment (4), association (8), assumption (7), attribute (2), being (1), biotic_community (1) ('community' in OED), causal_agency (1) ('cause' in OED), chance (4), citizenry (1), cleavage (1), cognition (1), collection (1), community (8), condition (1), condition (2), conflict (4), consequence (1), damnation (2), death (6), degree (2), dependency (1), disorder (3), distribution (3). **Hypernym level 2:** abandonment (1), abeyance (1), ability (3), abnormalcy (1) ('abnormality' in OED), absolution (1), absorption (6), acapnia (1), accident (2), accompaniment (1), account (1), achievement (1), acme (1), acquaintance (2), action (5), actuality (1), actus_reus (1), addiction (1), address (3), adeptness (1), adjudication (1), adolescence (2), adroitness (1), aestivation (2), affect (1), affection (1), affinity (5), affirmation (2), aftereffect (2), aftermath (2), agalactia (1) ('agalaxy' in OED). **Hypernym level 3:** abandon (2), abatement (1), abelian_group (1), aberrancy (1), abience (1), abiogenesis (1), abnegation (1), abortion (2), aboulia (1), abuse (3), abutment (1), acardia (1), accelerator (3), acceptance (1), acceptor (1), accession (1), accession (4), accommodation (5), accord (2), achylia (1), acicula (1), acidification (1), aclinic (1), acquaintance (1), acquirement (1), acrophony (1), actinism (1), adapid (1), adaptation (2). **Hypernym level 4:** abashment (1), abduction (2), aberration (2), abhorrence (1), abjuration (1), ablactation (1), accent (2), acceptability (1), accession (2), accession (6), accessory (2), acclimation (1), accommodation (6), accordance (2), accouchement (1), accretion (2), acetal (1), achondrite (1), achromatism (1), acid_anhydrides (1), acquired_taste (1) (see "acquired", ppl. A. in OED), action (3), action (9), acuteness (1), acyl (1), adamance (1), adaptability (1), add-on (1), addiction (2), addison's disease (1) (see 'addison' in OED). **Hypernym level 5:** 3-d (2) (see 'three' in OED), abandon (1), abandonment (3), abasement (2), abbacy (1), abdominousness (1), abiotrophy (1), ablation (1), ablation (2), abo_blood_group_system (1) ('blood group' in OED), abode (1), abomination (1), abortion (1), about-face (1), abscess (1), abseil (1), absolute_zero (1) (see 'zero' in OED), absolutism (3), absorption (5), abstainer (2), abstract (1), ac (2) (see 'alternating' in OED), acanthopterygii (1), acaricide (1), acceleration (2), acceptance (2), access (3), accident (1), accipiter (1). **Hypernym level 6:** abandonment (2) (see 'abandon' verb in OED), abarticulation (1), abasement (1), abattoir (1), abbreviation (1), abdicator (1), abdomen (2), aberrant (1) (see the adj in OED), abhorrer (1), abkhaz (1) (see 'abkhasian' adj, A, in OED), ablative (1), ablaut (1), abode (2), abominator (1), aborigine (1) ('aborigines' in OED), abrasion (3), ache (1), acid_test (1) (see 'acid' in OED), aconite (1), acoustic (1), acrimony (1), acroclinium (1), acronym (1), act (4), actomyosin (1), aculeus (1), acumen (1), ad (1), adactylia (1) (see 'adactylous' in OED), adansonia (1). **Hypernym level 7:** aa (1), abacus (1), abasia (1), abatis (1), abduction (1), abecedarian (1), aberdeen (1), abetment (1), abkhaz (1), abolitionist (1), abomasum (1), aborigine (2), abrader (1), abraham's bosom (1) (see 'bosom' in OED), abreaction (1), abridgement (1), abscondment (1) (see 'absconding' verb-form noun, in OED), absinth (1), absolutist (1), abstract (2), abstract art (1) (see 'abstract' in OED), abutment (2), abyss (1), academic degree (1) (see 'degree' in OED), acanthocephalan (1), acanthocyte (1), acanthosis (1), accent (3), acceptation (2), access (1). **Hypernym level 8:** aaron's_rod (1), abaca (2), abbess (1), acacia (1), academic (1), acarine (1), accelerator (4), accessary (1), abalone (1), abductor (1), abductor (2), abnegator (1), abortus (1), abseiler (1) ('abseil' in OED, "a person who descends a steep..."), absolute_pitch (1), abstraction (3), academic (1), acarine (1), accelerator (4), accentuation (1), acceptor (2), accessary (1), accidental (1), accommodation (2), accompaniment (2), accordionist (1), accusal (1), accused (1), ace (5), achilles tendon (1) (see 'tendon' in OED). **Hypernym level 9:** 1-hitter (1), 4wd (2), aardvark (1), aba (1), abbe (1), abbe_condenser (1) (see second'abbe' in OED), abbreviator (1), abetter (1), ablative_absolute (1), abnaki (1), abney_level (1) (see 'abney' in OED), abominable_snowman (1), absconder (1), absolute_ceiling (1) (see 'ceiling' in OED), abstemiousness (1), accelerator (1), accent (5), acciaccatura (1), accord (3), accordion (1), accoucheuse (1), accumulator (2), achimenes (1), acinus (1), ack-ack (1), acorn squash (1) (see 'acorn' in OED), actor's line (1), adducer (1), adjutant (1). **Hypernym level 10:** a_la_carte (1), abducens (1), abele (1), abortionist (1), absinthe (1), absolute_scale (1) (see 'absolute'), acarid (1), accentor (1), accidence (1), accipiter_cooperii (1), accordian_door (1), acridid (1), action (7), adenocarcinoma (1), adjunct (3), adonic (1), adventism (1), adz (1), aeolian harp (1) (see 'aeolian' in OED), aerogenerator (1) (see 'aero' in OED), aeroplane (1), affirmative_action (1), afghan (4), aflatoxin (1), afrikaans (1), agent_provocateur (1), agony_aunt (1), agony_column (1), agouti (1), aircraftman (1) (see 'aircraft' in OED). **Hypernym level 11:** abbey (2), acanthopterygian (1), accoucheur (1), adder (3), adjutant (1), admiral (2), african_elephant (1) (see 'african' in OED), agamid (1), ai (1), aircraft_carrier (1) (see 'aircraft' in OED), algorism (1), almanac (1), almond (1), amadavat (1), amazon_ant (1), ambulance (1), amphibian (2), amphibrach (1), amputator (1), anaesthetist (1), anapaest (1), angledozer (1) (see 'angle' in OED), anglicanism (1), anglo-french (1), angwantibo (1), anhinga (1), anorak (1), ant_bear (1), ant_cow (1) (see 'ant' in OED), ant_thrush (1). **Hypernym level 12:** aardwolf (1), adelie (1), admiral (1), agama (1), alienist (1), alpaca (3), alpine_fir (1) (see 'alpine' in OED), analyst (3), achovy (2), andean_condor (1) (see 'condor' in OED), angel_shark (1) (see 'angel' in OED, additions 1993), angler (3), anglo-catholicism (1), anole (1), arctic_fox (1), argentine (1), argus (2), asp (2), ass (3), babirusa (1), baboon (1), bactrian_camel (1) (see 'bactrian' in OED), baleen_whale (1) (see 'baleen' in OED), baltimore_bird (1) (see 'baltimore' in OED), barnacle (2), barracuda (1), basenji (1), basilisk (3), battle-ax (1), battle_cruiser (1). **Hypernym level 13:** abyssinian (1), african_green_monkey (1) (see 'green' in OED), alewife (2), alley_cat (1) (see 'alley' in OED), allosaur (1), american_elk (1) (see 'elk' in OED), american_redstart (1) (see 'redstart' in OED), antelope (1), baedeker (2), basking_shark (1) (see 'basking' in OED), bass (8), bison (1), blackcap (2), blennioid (1), blowfish (2), bluefish (1), bombard (1), bovine (1), bowhead (1) (see 'bow-head' in OED), boxer (4), brocket (1), brontosaur (1), bulldog (1), bullhead (2), cachalot (1), capelin (1), carangid (1), caribou (1), carrier_pigeon (1). **Hypernym level 14:** addax (1), afghan (5), airedale (1), alaskan_malamute (1) (see 'malamute' in OED), allice (1), alsatian (1), american_bison (1), american_plaice (1) (see 'plaice' in OED), anoa (1), appaloosa (1), arab (2), argal (1) (see 'argali' in OED), armed_bullhead (1) (see 'bullhead' in OED), bad_lands (1), basset (1), beagle (1), beaugregory (1), beluga (2), billy_goat (1), blackfish (1), blenny (1), bloodhound (1), bobcat (1), bongo (2), bonito (2), bonobo (1), bottlenose (2), briard (1). **Hypernym level 15:** affenpinscher (1), africander (1), albacore (1), amberfish (1), angora (1), aurochs (1), banteng (1), beef (1), bellwether (2), bezoar_goat (1) (see 'bezoar' in OED), bighorn (1), black_bass (1), blenheim_spaniel (1) (see 'blenheim' in OED), bronco (1), bull (1), bullock (2), bushbuck (1), caracul (1), cart_horse (1), charger (1), cheviot (1), cigarfish (1) (see 'cigar' in OED), cimarron (2), clumber (1), clydesdale (1), coach_horse (1), cocker (1), coohdog (1), cow (1), devon (2). **Hypernym level 16:** aberdeen_angus (1), ayrshire (1), beefalo (1), bullock (1), cero (1), chow_chow (1), durham (1), friesian (1), galloway (1), gaur (1), gayal (1), heifer (1), hereford (1), hind (1), jewfish (2), king_mackerel (1) (see 'king' in OED), santa_gertrudis (1), springer (2), texas_longhorn (1), whiteface (1).

References

Apresjan, J. (2000). *Systematic lexicography [Trans. Kevin Windle]*. Oxford: Oxford University Press.

Calzolari, N. (1988). The dictionary and the thesaurus can be combined. In *Relational models of the dictionary* (pp. 75–111). Cambridge: Cambridge University Press.

Changizi, M. A. (2001). Universal scaling laws for hierarchical complexity in languages, organisms, behaviors and other combinatorial systems. *Journal of Theoretical Biology, 211*, 277–295.

Changizi, M. A. (2003). *The brain from 25,000 feet: High level explorations of brain complexity, perception, induction and vagueness*. Dordrecht: Kluwer Academic.

Changizi, M. A. (2006). The optimal primate ventral stream from estimates of the complexity of visual objects. *Biological Cybernetics, 94*, 415–426.

Changizi, M. A. (2009). *X-ray vision and other superpowers you didn't know you have*. Benbella Books.

Changizi, M. A., & Shimojo, S. (2005). Character complexity and redundancy in writing systems over human history. *Proceedings of the Royal Society of London B, 272*, 267–275.

Changizi, MA., Zhang, Q., Ye, H., & Shimojo, S. (2006). The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. *The American Naturalist, 167*, E117–E139.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge: MIT Press.

Fellbaum, C., Grabowski, J., & Landes, S. (1998). Performance and confidence in a semantic annotation task. In C. Fellbaum (Ed.), *WordNet: An electronic lexical database* (pp. 217–238). Cambridge: MIT Press.

Fillmore, C. (1975) An alternative to checklist theories of meaning. In *Proceedings of the first annual meeting of the Berkeley linguistics society*. Berkeley: Berkeley Linguistics Society.

Goddard, C., & Wierzbicka, A. (Eds.). (2002). *Meaning and universal grammar—Theory and empirical findings* (Vol. I). Amsterdam: John Benjamins.

Goddard, C. (2006). Natural semantic metalanguage. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., pp. 544–551). Elsevier.

Goddard, C. (2007). Semantic molecules. In Mushin, Ilana & Mary Laughren (Eds.), *Selected papers from the 2006 annual meeting of the Australian linguistic society*. Brisbane, 7–9 July, 2006.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science, 4*, 195–208.

Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. Chicago: University of Chicago Press.

Moldovan, D., & Novischi, A. (2004). Word sense disambiguation of WordNet glosses. *Computer Speech and Language, 18*, 301–317.

Ravasz, E., & Barabasi, A.-L. (2003). Hierarchical organization in complex networks. *Physical Review E, 67*, 026112.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). New Jersey: Lawrence Erlbaum.

Sigman, M., & Cecchi, G. A. (2002). Global organization of the WordNet dictionary. *Proceedings of the National Academy of Science, 99*, 1742–1747.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29.

A, Wierzbicka (1996). *Semantics: Primes and universals*. New York: Oxford University Press.

Wierzbicka, A. (in press). The theory of the mental lexicon. HSK (Handbücher zur Sprach- und Kommunikationswissenschaft. Handbook of Linguistics and Communication Science.) Slavic Languages (Mouton: Walter de Gruyter, Berlin/New York).