

## Lexical Patterns: From Hornby to Hunston and Beyond

Patrick Hanks  
Masaryk University

*I start with a brief summary of A. S. Hornby's achievement in creating the Idiomatic and Syntactic English Dictionary (ISED 1942), a work which gradually mutated, through many editions, into the present Oxford Advanced Learner's Dictionary of Current English. Among Hornby's radical innovations was a focus on examining the patterned nature of language and presenting patterns of word use in a succinct form for assimilation by language learners. He saw that each verb is associated with a different set of syntactic patterns, and he was able to impose order on apparent chaos by picking out structural threads and establishing templates for pattern analyses. The 5th edition of OALD, edited by Jonathan Crowther (1995) and the 6th edition, edited by Sally Wehmeier (2000) were recensions of Hornby's work using corpus evidence.*

*Hornby and his mentor, H. E. Palmer, had an intuitive understanding of the patterned nature of language, but they lacked the evidence that was necessary for a detailed empirical study of the collocational patterns associated with different meanings of each word. This had to wait until the advent of very large corpora, inspired in particular by the vision and practice of J. M. Sinclair. As early as 1966, Sinclair predicted that patterns of lexis "would not yield to anything less than a very large computer". Much of his life's work was devoted to developing sound linguistic theory on the basis of empirical analysis of corpus evidence. His principles were taken up by subsequent linguists, for example Alan Partington, Michael Hoey, Susan Hunston, and Gill Francis.*

*In his 1987 paper entitled "The nature of the evidence", Sinclair stresses the importance of distinguishing significant collocations from random co-occurrences. The first attempt to undertake statistical analysis of collocations in a corpus was by Church and Hanks (1990), but it was not until Kilgarriff, Rychlý, and their colleagues developed the Word Sketch Engine (Kilgarriff et al. 2004) that a user-friendly tool was made widely available for people to see at a glance how the meanings of a semantically complex word are associated with and indeed activated by its collocates.*

*Modern corpus tools such as these bring us full circle, back to Hornby's original vision of patterns of word use and word meaning. It is now possible to examine that vision in the light of massive bodies of evidence. Not only does this lead inexorably to new theoretical insights into the nature of language, it also make it possible to develop new kinds of dictionaries for human learners and computational applications alike—dictionaries that focus rigorously on patterns of word use, rather than (say) on historical semantics and morphology.*

*In the second part of the lecture, I give a progress report on the corpus-driven Dictionary of English Verb Patterns currently being developed at the Masaryk University in Brno. I compare the Pattern Dictionary with Pattern Grammar and discuss some problems of lexicographical analysis, such as finding the right level of generalization for each element in a pattern. How is one sense to be distinguished from another? For a word with many patterns, are some patterns more important than others, and if so why? How are creative uses of a word distinguished from more mundane uses? What is the role, in pattern analysis, of an ontology?*

### **1. A. S. Hornby and English lexicography in the 1930s**

In 1923 a shy young man of 25 called Albert Sydney Hornby (known affectionately to his friends and colleagues as “Ash”), armed with a degree in English from University College, London, sailed to far-away Japan to start a career as a teacher of English. This event was to have far-reaching consequences for English lexicography. Hornby proved to be a gifted and skilled teacher, who had sound theoretical instincts and a motivation to explain the meaning and use of English words in terms that ordinary students could understand and assimilate. In 1931 he was invited by Harold E. Palmer, director of the Tokyo Institute for Research into English Teaching, to participate in a programme of vocabulary research. Five years later, in 1936, when Palmer left Japan, Hornby was appointed head of research at the Institute.

At the institute in Tokyo, Hornby compiled lists of important collocations in English, using his wide reading and his intuitions as a teacher of English. He worked with Palmer on English verb syntax and on vocabulary selection for learners at different levels. At least three of the insights of Palmer, Hornby, and their colleagues in the 1930s have provided a principled foundation for much subsequent work, including modern corpus-driven lexicography (which, it should be said, is still in its infancy today). These three principles may be summarized as follows:

1. Language in use is highly patterned. Each word is typically associated with only a small number of syntactic patterns.
2. Ordinary everyday communication consists of utterances based on patterns of usage built up around a small number of very frequent words, each of which is used in a comparatively small number of patterns or structures. At the same time, usage also encompasses a very large number of other possible and actual words and structures, some of which are used only very rarely.
3. The verb is the pivot of the clause. In the front matter of OALD, Hornby asserts: “Verb patterns are the most important”, and urges learners to “spend a few hours studying ... verb patterns”, as “the ordinary grammar-book and dictionary usually fail to supply adequate information on such points.”

Hornby and his colleagues were well aware that the English monolingual dictionaries available in the 1920s and 30s did not take account of these principles. In fact, those dictionaries were quite unsuitable for pedagogical purposes. The focus in those days—for example in Fowler’s brilliant *Concise Oxford Dictionary* (COD) of 1911, the best-selling English dictionary of its time, between the wars—was on historical philology. From its first publication, the full title of COD was *The Concise Oxford Dictionary of Current English*, but it was not until the 8th edition (1990, edited by Robert Allen) that COD really earned that subtitle, and even then it gave no account of structured patterns of usage. In the first seven editions of COD, prominence was given to word history and etymology, despite the subtitle. The oldest known meaning of each word was placed first, provided only that it was still current<sup>1</sup>. For example, the noun *carnation*, denoting a kind of sweet-smelling flower, was nested under the adjective *carnal*, meaning ‘of or pertaining to flesh’, because both were thought to be derived from Latin *carne* ‘flesh’. In those days, etymologists thought that the flower was so named because of its fleshy pink colour.<sup>2</sup> The first sense of the noun *camera* was given as ‘a small vaulted room’, not an apparatus for taking photographs, even though photography was already a well-established technology in Fowler’s day and the ‘small room’ sense of *camera* was already rare or obsolete. This quaint editorial policy, inherited from the great historical lexicographical enterprises of the 18th and 19th centuries, was and is associated with some curious value judgements about the nature of language and word meaning, for example:

1. that older meanings are somehow better than modern meanings,
2. that the language of our forebears is somehow better than our own, and
3. that “the language (whatever language it may be: English, French, Spanish, Catalan, Latin, Greek, or other) is going to the dogs”.

The belief that the language is going to the dogs has been around at least since the 5th century BC (it was satirized in ribald terms by Aristophanes), and it continues to be reflected in at least two of America’s best-selling dictionaries, published under the name of America’s first and most belligerent lexicographer, Noah Webster.

As a practical teacher of English, Hornby recognized that historical principles of lexicography are irrelevant to effective language learning and that learners need a dictionary offering practical rules and models of current usage on which to build their own competence. Unlike many of his contemporaries, he decided to do something about it. With the aid of two colleagues, Edward Gatenby and Harold Wakefield, he set about compiling an *Idiomatic and Syntactic English Dictionary*

---

<sup>1</sup> An exception to the historical order of senses in early editions of COD was that senses that had become totally obsolete were relegated to last position and labelled “Obs.”. This was no doubt the justification for the subtitle.

<sup>2</sup> Modern etymologists think that *carnation* is more probably an alteration, by folk etymology, of Arabic *qaranful* ‘clove or clove pink’, from Greek *karyophyllon*.

(ISED). This was completed in 1941 and published by Kaitakusha in 1942.<sup>3</sup> It was the first dictionary of English as a foreign language, initiating a genre that has evolved into a rich variety of present-day forms. In 1948, it was re-published unaltered by Oxford University Press under the new title *A Learner's Dictionary of English*. It became an international and perennial best-seller. In the third edition (1974), the name *Oxford* was added to the title.

One of the pleasures of preparing this lecture was revisiting the first edition of Hornby, Gatenby, and Wakefield and discovering for how fresh, clear, readable, and easy to understand that first edition was. In the second (1963) and subsequent editions, Hornby lost some of that freshness and ease of use, apparently under the influence of the then current 4th edition (1951) of the *Concise Oxford Dictionary*. He made the following changes among others:

- Putting in many thousands of additional entries and subentries, greatly increasing coverage.
- Nesting subentries under root words, e.g. he moved *blackbird* and *blackboard* and nested them under *black*.
- Using a swung dash to represent repetition of the headword within an entry, so that *blackbird* and *blackboard* are represented as *~bird* and *~board*.
- Rewriting definitions in a more formal style, apparently in pursuit of the principles of consistency and substitutability. Many of ISED's glossed examples became formal definitions. So, for example, ISED had glossed examples like this, one of several under *blame*:
  - o *He blamed his failure on the teacher [blamed the teacher for his failure]*  
(= he said that it was the teacher's fault)

In OALD2 this was swept away and subsumed with other examples under a general definition of the verb:

- o find fault with; fix the responsibility on (sb. or sth.) (*for sth.*): *Bad workmen often ~ their tools. He ~d the teacher for his failure.* (Colloq.)  
*He ~d the teacher for his failure.*

It is not clear that these 1963 changes were entirely beneficial. The increase in coverage undoubtedly gave the dictionary more potential usefulness as an aid for decoding tasks (i.e. for reading and understanding), but it reduced its usefulness as an encoding aid (for writing and speaking idiomatically) by making it harder for

---

<sup>3</sup> The first edition is still available on Kaitakusha's website, with a blurb written over 66 years ago: "This dictionary has been compiled to meet the needs of foreign students of English. It is called *Idiomatic and Syntactic* because the compilers have made it their aim to give as much useful information as possible concerning idioms and syntax. It is hoped that the dictionary will be of value to those who are learning English as a foreign language." The publisher might care to note in some future version of this blurb that, during the intervening two-thirds of a century, those hopes have been amply fulfilled.

learners to find what they are looking for. This adverse effect was compounded by the policy of nesting and the use of the swung dash. The swung dash undoubtedly saved some space, but it made words much harder to recognize and may well have baffled some learners. At any rate, thirty years later, in the 5th edition (1995), OALD abandoned the swung dash, and in the 6th edition (2000) the policy of nesting was likewise abandoned: *blackbird* and *blackboard*, along with thousands of other compounds, were restored to the headword status that Hornby and co. had originally given them.

In 1942, shortly after the outbreak of war and shortly before publication of their dictionary, the three lexicographers left Japan under a programme for the exchange of enemy nationals. Both Hornby and Gatenby went on to distinguished careers in the British Council.

In 1954, Hornby published *A Guide to Patterns and Usage in English*, a lexically based practical and partial grammar whose approach to analysis and tabular style of presentation reflected the methods employed in the *Advanced Learner's Dictionary*.

The *Advanced Learner's Dictionary* went through several editions under the editorship of Hornby and, subsequently, some other able lexicographers, including Tony Cowie and Jonathan Crowther. A much appreciated feature of the early editions was the guidance given on grammar and usage, on principles that had been devised by Hornby and Palmer. The 5th edition, edited by Jonathan Crowther (1995) and the 6th edition, edited by Sally Wehmeier (2000), were radical recensions of the work of Hornby and other previous editors in the light of corpus evidence. Hornby's name continues to grace the title page of the current (7th) edition of OALD, although his co-workers have been consigned to oblivion. Hornby himself acknowledged what most lexicographers know but the public perhaps do not, namely that lexicography is a team game. He was always careful to pay tribute to the work of his original collaborators (Gatenby and Wakefield) and their successors.

## 2. Clause roles, Hornby's Verb Patterns, and OALD

Hornby's recognition of the patterned nature of usage and the central importance for language learners of knowing the syntagmatics of verbs led him to formulate a summary of English verb patterns. He frequently drew attention to the danger for foreign learners of false analogy, for example, forming an ungrammatical sentence such as “\*I proposed him to come”, either by false analogy with a verb pattern in their own native language or by false analogy with well-formed grammatical sentences in English such as “I asked him to come” and “I told him to come”. It is important, therefore, for learners to know, not only what verbs means, but also how to use them. Hornby believed that this could be achieved if learners would spend a few hours memorizing the verb patterns, so that, for example, before using the verb *propose* in an essay, they could look it up in OALD, see that it is used in verb pattern 9, and thereby know that the correct idiomatic phraseology is “I proposed that he should come”, not “\*I proposed him to come”.

This was done systematically for each verb in the dictionary, by stating a pattern number in square brackets alongside definitions and examples, e.g. [VP9]. The pattern numbers refer to a look-up table in the front matter of the dictionary. In the 1974 edition of ALD, for example, VP9 is “S + *vt* + *that*-clause” (as in “*I proposed that he should come*”). This contrasts with VP10, “S + *vt* + dependent clause/question” (as in “*She asked whether he would come*”) and VP11, “S + *vt* + noun/pronoun+ *that*-clause” (as in “*I told him that he should come*”).

It would difficult to understate the importance of this insight from the point of view of lexical and grammatical theory. As we shall see, it plays a central role in corpus-based research into the relationship between meaning and use.

A word is needed here on clause roles. In 1963 Hornby did not consider the subject of the clause to be part of the pattern. In 1974, he and Tony Cowie did. This, in my opinion, was a step in the right direction, from which, unfortunately, EFL lexicographers have since backed away. Its importance only becomes apparent when an attempt is made to assign semantic values to clause roles, in order to distinguish one sense of a verb from another. The subject of a clause is part of its pattern. For the vast majority of clauses, the subject has the default semantic value [[Human]]. These constitute the unmarked cases. More interesting are the marked cases, where the semantic value of the subject is not [[Human]]. Examples are: *one administrative entity swallowing up another administrative entity* (distinct from *a human swallowing a physical object*), or *an ideology firing people with enthusiasm* (distinct from *a human firing people from their jobs*). The semantic relationship between a verb and the rest of the clause is a relationship among clause roles, not merely between the verb and various nouns, adjectives, or prepositions. This is a fine point, but failure to take note of it has sometimes led to confusion in lexical analysis.

The terminology of generative linguistics, which makes binary divisions and refers to the subject as the “external argument”, lumping adverbials and objects together as part of the “verb phrase” (which, in more esoteric terminology, is sometimes called the “inflection phrase”), is unhelpful in this respect. Empirical linguists such as Quirk and (with minor variations) Biber, Sinclair, Halliday and others, recognize five clause roles, in an analytic structure which has come to be known informally by the mnemonic SPOCA. Since grammar these days is a hotbed of terminological confusion, with considerable potential for misunderstanding, it is worth taking a few moments to summarize the five basic clause roles. They have a central part to play in verb pattern analysis. They are:

- |   |            |   |
|---|------------|---|
| S | Subject    | Obligatory in English, except in a) imperatives (e.g. “ <i>eat your greens!</i> ”) and b) elliptical clauses such as natural responses, e.g. “ <i>What was he doing?</i> ” — “ <i>eating an apple</i> ” |
| P | Predicator | The verb group, including auxiliary verbs and negatives, but not nouns or prepositional phrases. By some writers, V (verb) is used in place of P.   |

- O Object English clauses may have one, two, or no objects. In SPOCA, the object is a clause role in its own right, not part of the verb phrase, as it is in grammars based on predicate logic
- C Complement A clause role that is co-referential with either the subject or the object of the clause. Examples of subject complements are the adjective *happy* in *he seems happy* and the noun phrase *the pronunciation editor* in *Carolyn was the pronunciation editor*. Examples of object complements are as in *She made him happy; they appointed her pronunciation editor*.
- A Adverbial (sometimes called Adjunct). A clause may have any number of adverbials. A distinction is made between obligatory adverbials (for example, the locative adverbial *on the table* in *He put the cup on the table*) and optional adverbials (for example, the time adverbial in *He died in 1974*).

Translated into SPOCA, Hornby's 1963 patterns look like this:

1. S P O. *We lit a fire.*
2. S P {to/INF}. *He wants to go.*
3. S O {to/INF}. *They want him to go.*
4. S P O (to be) SC. *I consider it (to be) a shame.*
5. S P INF. *I made him do it. Will you help me carry this box?*
6. S P O {V-ing}. *He kept me waiting. I saw him running off.*
7. S P O OC (adj.). *Don't get your clothes dirty.*
8. S P O OC (noun). *They elected him president.*
9. S P O {V-en}. *He got the document printed. I have never heard Italian spoken.*
10. S P O A. *He took his hat off. Mr Smith showed me to the door.*
11. S P {(that) CL}. *I suppose (that) he will be late.*
12. S P O {(that) CL}. *I warned you that he would be late.*
13. S P {Wh- CL}. *I know why he did it.*
14. S P O {Wh- to/INF}. *We showed him how to do it. They told him when to start.*
15. S P {Wh- CL}. *I wonder what it is. I don't mind where we go.*
16. S P O {Wh- CL}. *Tell me what it is. Ask him where he put it.*
17. S P {V-ing}. (A) *They stopped talking.* [Compare *They stopped to talk*—different meaning] (B) *He began talking.* [Compare *He began to talk*—same meaning] (C) *It needs doing.* [Compare *it needs to be done*—passive meaning]

- 
18. S P O A. (A: with *to*, alternating with 19): *He gave some money to his wife.* (B: with *for*, alternating with 19): *He bought a watch for his wife.* (C, with other prepositions, not alternating with 19): *They criticized him for being late. He was throwing stones at a dog.*
  19. S P O O. (A): *He gave his wife some money.* (B): *He bought his wife a watch.* (C): *Lord, forgive us our sins. The rain lasted all day.*<sup>4</sup>
  20. S P C [C expressing duration, distance, price, or weight]. *It lasted all day. We walked (for) five miles. His car cost €12,000. It weighs five tons.*
  21. S P. *Birds fly. We all eat, breathe, drink, and die. The sun was shining.*
  22. S P C. *This is a book. This book is mine. The leaves have turned red.*
  23. S P A. *The sun rises in the east.*
  24. S P A. *He called on me.*
  25. S P {*to*/INF}. *We stopped to have a rest.*

Admiration for Hornby's insights into the nature of syntax and his organized presentation of pattern structures should not blind us to the fact that there are some problems with the way that he presented the data. In the first place, there are 25 verb patterns, which is a lot for a learner to memorize and know how to apply. This is made harder by a number of subtle semantically motivated subdivisions and by the fact that the clause roles are expressed in abstract terms, being referred to by numbered references to a look-up table, rather than by a phrase or name with mnemonic value.

A second problem is that not only did Hornby revise his patterns from time to time, but also he changed their order and numbering. It must have been hard for teachers and students who had taken the trouble to memorize the patterns of the 1963 edition to relate them to the new order and recognize that, say, VP11 of 1963 corresponded to VP9 of 1974. There is no obvious way of associating the term "VP9" or "VP11" with a *that*-clause. Conscientious teachers and learners must have spent many hours thumbing back to the front matter of the dictionary. Less conscientious users would have simply ignored them, thus failing to benefit from the important information about idiomatic phraseology which they encapsulated.

A third problem is that in Hornby's work there is no obvious motivation for the order of patterns as they are presented in the front matter of the dictionary—no differentiation, for example, between simple clause structures on the one hand and more complex structures involving subordinate clauses or infinitive forms on the

---

<sup>4</sup> Subclasses A and B of pattern 19 alternate with pattern 18 (SPOA)—e.g. *He gave some money to his wife; he bought a watch for his wife.* Subclass C does not so alternate. In the second example of C, the phrase *all day* is not really an object at all, but rather a time adverbial, even though it does not have a prepositional head.

other hand<sup>5</sup>. They are all jumbled together, higgledy-piggledy, and there is quite a lot of overlap. There are some over-subtle distinctions.

These are no doubt among the reasons why the verb patterns were eventually greatly simplified in more recent editions of the *Oxford Advanced Learner's Dictionary*. By the time of Crowther's 5th edition (2000), patterns were no longer identified by numbers but rather by abbreviated phrases with mnemonic value. Thus the 1974 "VP11", with its rather clumsy front-matter apparatus "S + *vt* + noun/pronoun+ *that*-clause", had become a simple mnemonic: "Vn (that)". Here, each element in the pattern name has mnemonic value: V means 'verb', n means 'noun', and '(that)' signifies a clausal complement introduced by the subordinating conjunction *that*. The conjunction is often omitted in informal English speech and writing, hence the brackets.

In OALD6, the patterns are set out more clearly in the front matter, but in a much reduced form. The emphasis is on streamlining the presentation for the user. The number of patterns has been reduced from 25 to 20, but there is little significant loss of information. Pattern numbers have been abandoned in favour of mnemonics, and the technical grammatical terminology has been reduced to a minimum. In the front matter, the summary of patterns is more carefully ordered. Patterns that take clauses are separated from the rest: the 20 patterns are summarized and organized under six subheadings, in order of gradually increasing complexity, with example sentences, as follows:

Intransitive verbs

[V]                    *A large dog appeared.*

[V + *adv/prep*]    *A group of swans floated by.*

Transitive verbs

[VN]                    *Jill's behaviour annoyed me.*

[VN + *adv/prep*]   *He kicked the ball into the net.*

Transitive verbs + two objects

[VNN]                 *I gave Sue a book for Christmas.*

<sup>5</sup> A difficulty related to this last point is that the learner had to deal with two types of grammatical element: the functional SPOCA elements that encode 'who is doing what to whom' in a clause, and formal elements such as *to/INF*. For example, Hornby's pattern 3 is "SVO *to/INF*". Some modern descriptive linguists would see this as involving two different subtypes of O. For instance, Francis et al. (1996) identify Hornby's pattern 3 as an SVOO pattern, with *to-infinitive* being regarded as an object on a different syntactic 'layer', thus:

	Verb group	Noun group	<i>to</i> -infinitive clause
Subject	Verb	Object	Object
My girlfriend	nagged	me	to cut my hair.

## Linking verbs

[V-ADJ]	<i>His voice sounds hoarse.</i>
[V-N]	<i>Elena became a doctor.</i>
[VN-ADJ]	<i>She considered herself lucky.</i>
[VN-N]	<i>They elected him president.</i>

## Verbs used with clauses or phrases

[V <b>that</b> ]	
[V ( <b>that</b> )]	<i>He said that he would prefer to walk.</i>
[VN <b>that</b> ]	
[VN ( <b>that</b> )]	<i>Can you remind me that I need to buy some milk?</i>
[V <b>wh-</b> ]	<i>I wonder what the job will be like.</i>
[VN <b>wh-</b> ]	<i>I asked him where the hall was.</i>
[V <b>to</b> ]	<i>The goldfish need to be fed.</i>
[VN <b>to</b> ]	<i>He was forced to leave the keys.</i>
[VN inf]	<i>Did you hear the phone ring?</i>
[V <b>-ing</b> ]	<i>She never stops talking.</i>
[VN <b>-ing</b> ]	<i>His comments set me thinking.</i>

## Verbs + direct speech

[V speech]	<i>“It’s snowing,” she said.</i>
[VN speech]	<i>“Tom’s coming to lunch,” she told him.</i>

In the dictionary itself, if a sense of a verb participates in more than one pattern, the patterns are stated alongside individual examples, rather than before the definition. Thus, sense 1 of *propose*, “to suggest a plan, an idea, etc., for people to think about and decide on”, is illustrated by no less than seven example sentences, showing participation in no less than five patterns, one of which records a British/American variation in the wording of the subjunctive in the *that* clause:

◇ [VN] *The government proposed changes to the voting system.* ◇ *What would you propose?*

◇ [V **that**] *She proposed that the book be banned.* ◇ (BrE also) *She proposed that the book should be banned.*

◇ [VN **that**] *It was proposed that the president be elected for a period of two years.*

◇ [V **-ing**] *He proposed changing the name of the company.*

◇ [VN **to** inf] *It was proposed to pay the money from public funds.*

Let us come back to Hornby's original point, namely that “\**I proposed him to come*” is a grammatical error. Is this contradicted by the third and fifth patterns in this entry? OALD6 tries to explain in a “help note”: “This pattern is only used in the passive.” Unfortunately, it does not say that this comment applies also to the “[VN **that**]” pattern. Moreover, “VN” is, as a matter of fact, never true of the verb *propose* in this sense. You can *propose a plan or idea*, but you cannot \**propose a person to do something* or \**propose a person that something*. The impersonal passive, with proleptic *it*, cannot be equated with an object of an equivalent active use. You cannot say “\**The government proposed him to pay the money from public funds*” or “\**We proposed the government to pay the money from public funds.*” Hornby was right all along; by oversimplifying the grammatical apparatus, his successors have got it wrong with regard to the impersonal passive, which must be recognized as a pattern in its own right, not treated as a transformation of an active.

I mention this small point to illustrate just how difficult it is to get the details of patterns of idiomatic usage right. Grammatical descriptions as subtle, detailed, and factually accurate as that of OALD6 can only be teased out, word by word, with results achieving reasonable accuracy, by painstaking analysis of large quantities of corpus data, supported by a reasonably sophisticated grammatical apparatus.

An excellent research topic for a Ph.D. dissertation would be a comparison of verbs and verb patterns in the pre-corpus 1963, 1974, and 1989 editions of Hornby's and Cowie's OALD with the corpus-based 5th edition, edited by Crowther and the 6th edition edited by Wehmeier. This would not merely be of historical interest. Because neither Hornby nor Cowie had a corpus—corpora had not been invented in their day—they were reliant to a large extent on their intuitions when describing patterns. No doubt these intuitions were excellent and finely tuned, as they were both experienced, insightful, and widely read teachers of English, but there are many places in the dictionary where corpus evidence have prompted their successors to revise their entries. A systematic comparison of the pre- and post-corpus editions, focusing in particular on the description of verb behaviour, would shed valuable light on the complementary roles of evidence and intuition in a succession of highly skilled lexicographical teams striving to achieve what is essentially the same goal, using very similar descriptive apparatus but very different kinds of source data.

### 3. Patterns in English Dictionaries since Hornby

It is instructive to compare the treatment of verb patterns in monolingual English dictionaries since Hornby. A certain amount of ambivalence on the part of lexicographers may be detected. Is it really the job of the dictionary to explain grammar in the tradition of Hornby, and if so, how should it be done? The general tendency in recent years has been for British dictionaries to focus on word meaning and word classes, to minimize the explicit grammatical apparatus and terminology, to describe collocates in terms of their word classes rather than their clause roles or semantic types, and to attempt to convey grammatical information by judicious selection (or in OALD's case, construction) of examples.

Let us first look in more detail at the entry for *propose* in OALD6 and consider some issues of lexicographic principle. Although detailed discussion of a single verb does not constitute a statistically valid sample for purposes of dictionary evaluation, it does raise some interesting theoretical and practical points, which have far-reaching consequences. The full entry in OALD6 is as follows:

**propose**, *verb*

[SUGGEST PLAN] **1** (*formal*) to suggest a plan, an idea, etc., for people to think about and decide on: [VN] *The government proposed changes to the voting system.* ◇ *What would you propose?* ◇ [V **that**] *She proposed that the book be banned.* ◇ (BrE also) *She proposed that the book should be banned.* ◇ [VN **that**] *It was proposed that the president be elected for a period of two years.* ◇ [V **-ing**] *He proposed changing the name of the company.* ◇ [VN **to inf**] *It was proposed to pay the money from public funds.* [HELP] This pattern is only used in the passive.

[INTEND] **2** to intend to do sth: [V **to inf**] *What do you propose to do now?* ◇ [V **-ing**] *How do you propose getting home?*

[MARRIAGE] **3** ~ (sth) (to sb) to ask sb to marry you: [V] *He was afraid that if he proposed she might refuse.* ◇ [VN] *He proposed marriage.*

[AT FORMAL MEETING] **4** [VN] ~ sth | sb for /as sth to suggest sth at a formal meeting and ask people to vote on it: *I propose Tom Ellis for chairman.* ◇ **to propose a motion** (= to be the main speaker in support of an idea at a formal debate)—compare OPPOSE, SECOND.

[SUGGEST EXPLANATION] **5** [VN] (*formal*) to suggest an explanation of something for people to consider [SYN] PROPOUND: *She proposed a solution to the mystery.*

[IDM] **propose a toast (to sb) | propose sb's health** to ask people to wish sb health, happiness, and success, by raising their glasses and drinking.

With the exception of the quibble about the impersonal passive, discussed above, this is lexicography of the highest order of delicacy, clarity, and accuracy—about as good as it gets, given existing assumptions about lexicographic principles. The definitions are clear, the grammatical description is well thought out, and the example sentences are (with minor exceptions) well constructed for the benefit of learners.<sup>6</sup>

---

<sup>6</sup> Example sentences in OALD6 and 7 are designed to illustrate patterns of normal usage. Generally, they are not actual quotations from texts in a corpus, but rather corpus-inspired constructs designed to illustrate linguistic competence. In this respect, the recent editions of OALD differ from other corpus-based dictionaries. I shall say no more about this controversial issue here.

It is with some diffidence, therefore, that I will now suggest a move towards new lexicographical principles. My suggestions are inspired by corpus analysis. I would like to believe that, if Hornby had had access to corpus evidence, he would have been sympathetic to these proposals. The aim is to take the best in traditional and current lexicographical practice and ask whether it could be better. I propose eight new principles.

1. *Avoid fine-grained semantic distinctions*

Computational linguists often assert that distinctions in dictionary definitions are “too fine-grained”. One motive in this complaint is that computational linguists want definitions to be mutually exclusive, but this is a mistake. It confuses natural language with predicate logic. There is much overlap everywhere in matters of word meaning. Nevertheless, it may be that, as lexicographers, we have something to learn from this more general complaint: it can also be read as a polite way of telling us that some dictionary entries are not merely too fine-grained but needlessly repetitious. In OALD’s entry for *propose*, is it really necessary to make a distinction between senses 1 and 5, for example? Proposing an idea and proposing an explanation are semantically very close and should perhaps not be distinct, since it is simply not possible to come up with two distinct lexical sets of direct objects: one of nouns that mean ‘idea’ and the other of nouns that mean ‘explanation’. Consider the phrase ‘propose a hypothesis’. Is this sense 1 or sense 5 of *propose*? Is a *hypothesis* an idea or an explanation? The answer, of course, is that it is both. Sets of direct objects often constitute a chain of overlapping Wittgensteinian family resemblances. It is hard to defend the idea that there are two different transitive senses of *propose* in this case. A few sharp slashes of Ockham’s razor (avoiding the needless replication of entities) are called for. (I hasten to add that OALD is not the only dictionary that makes this unnecessary distinction.)

2. *Do not confuse domain with meaning*

*Proposing a motion* (sense 4) is semantically identical to *proposing a plan for people to think about and decide on* (sense 1). They belong together. OALD6 rightly gives information about the domain in sense 4 (“*at a formal meeting*”), but this does not need to be dressed up as a semantic distinction. A domain-commented example at sense 1 would be clearer and more elegant.

3. *Take account of the semantic types of collocates*

In contrast to the previous point (2), sense 4 needs to be split. *Proposing someone for/as a specific role* does not fit well semantically with *proposing a motion*. Both uses of this verb are indeed typical of the domain of formal meetings and parliamentary procedure, but there the similarity ends. There is a semantic difference. The two meanings (*propose a motion* and *propose*

*a person for a role*) are distinguished by the semantic types [[Person]] and [[Proposal]], and the explanation will be clearer if they are kept separate. A person is not a proposal.

4. *Patterns should play a role in organizing the entry*

The preceding point implies that patterns which have different meanings should be treated as different senses even if they are in the same domain. Carrying this idea a step further, we see that sense 1 of *propose* in OALD6 is associated with no less than five patterns. This is not a problem in itself, but the question must be asked: are they all in the right place? Consider the problematic last pattern, “◇ [VN to inf] *It was proposed to pay the money from public funds.*” Semantically, this is on a borderline between the [SUGGEST PLAN] sense and the [INTEND] sense. As so often happens, there is no sharp boundary between the two senses. For lexicographical purposes, however, a clearer focus will be achieved if all the [to inf] patterns are grouped at the [INTEND] sense, i.e. sense 2, not sense 1. This will work well because a *to*-infinitive governed by the verb **propose** always signals intended action.

5. *Seek the right level of delicacy in pattern description*

In sense 3, the [MARRIAGE] sense, the essential point is that, if there is no direct object, i.e. if *propose* is intransitive, the normal meaning is “ask someone to marry you”, not “suggest a plan to them”. The intransitive pattern is the normal one, and for the sake of clarity of exposition, it should be separated out, not integrated with other possible ways of realizing the same meaning. Conversely, some patterns crop up, rather confusingly, in several different senses of: for example, there are four occurrences of the verb pattern “[VN]” in this entry: in senses 1, 3, 4, and 5. Do these really represent different senses? Would anything be lost if they were lumped together? If they really represent different senses, can they be differentiated according to the semantic type of the nouns involved?

6. *The right level of pattern delicacy implies sorting according to semantic type, not just word class*

The addition of a [VN] pattern to sense 3, reinforced by the rubric “~ (sth) (to sb)”, is less than helpful. It muddies the waters: if taken literally, it could be read, wrongly, as implying that a sentence such as *John proposed a swim (or a cycle ride) to Mary* means that he asked her to marry him! The right level of delicacy requires explicit mention of the noun *marriage*, not the indefinite pronoun *sth* (*something*). This apparently simple and obvious point has far-reaching consequences.

For the vast majority of transitive uses of *propose*, the preferred semantic type of the direct object is [[Event]] or [[Plan]]. Unfortunately, this is obscured by another common linguistic phenomenon, namely ellipsis. In examples such as ‘Local government officials were able to **propose**

new dual-carriageway trunk roads’, the underlying meaning is ‘Local government officials were able to *propose the construction of new dual-carriageway trunk roads*’, where the absent noun *construction* is of semantic type [[Event | Plan]].

7. *Seek the right level of delicacy in syntactic comments*

The HELP note at the last pattern of sense 1 is misleading because it is underrestricted. This pattern is normally found only in the impersonal passive. It would, for example, be stretching idiomaticity to say \**Money was proposed to be paid from public funds*. This is exactly the sort of invented borderline example—just about possible, but bizarre or abnormal and not supported by evidence—that has bedevilled armchair linguistics for the past half century and led to much pointless speculation about a sharp dividing line between syntactically well-formed and ill-formed sentences.

A similar point arises regarding the  $\diamond$  [VN **that**] pattern, also in sense 1. *It was proposed that the president be elected for a period of two years* is likewise an impersonal passive. It is not idiomatic to say \**The president was proposed that he be elected for two years*, still less the active, which I suppose would have to be something like \**They proposed the president that he be elected for two years*, which is gibberish.

8. *The right level of syntactic delicacy implies clause role description, not word classes*

The apparatus of OALD6’s verb patterns is beautifully clear and simple. So at first sight it may appear unnecessarily pedantic to insist that VN ought to be SPO. In this case, it makes little difference, but elsewhere use of word classes in place of clause roles has led to errors in analysis, as we shall see shortly. It is also important to distinguish patterns that are typically active from those that are typically passive.

Similar points could be made about other verbs, and not only in OALD but also in all the other leading learner’s dictionaries, for they are all *meaning-driven*. What I am proposing here, in a nutshell, is a dictionary that is not merely *corpus-driven*, but *pattern-driven*.

To show how this works, let me start by illustrating a possible new version of the OALD entry, taking account of the above points:

**propose**, *verb*

[SUGGEST PLAN] 1 to suggest a plan, an idea, or an explanation of something, for people to think about and decide on: [S P O] *The government proposed changes to the voting system.*  $\diamond$  [S P O] *What would you propose?*  $\diamond$  [S P **that**] *She proposed that the book be banned.*  $\diamond$  (BrE also) *She proposed that the book should be banned.*  $\diamond$  [ it be P (impersonal passive) **that**] *It was proposed that the president be elected for a period of two years.*  $\diamond$  [S P

**-ing**] *He proposed changing the name of the company.* ◇ [AT A FORMAL MEETING] **to propose a motion** (= to be the main speaker in support of an idea at a formal debate)—compare OPPOSE, SECOND.

[INTEND] **2** to intend to do sth: [S P **to** inf] *What do you propose to do now?* ◇ [S P **-ing**] *How do you propose getting home?* ◇ [it be P (impersonal passive) **to** inf] *It was proposed to pay the money from public funds.*

[MARRIAGE] **3** [S P] to ask sb to marry you: *He was afraid that if he proposed she might refuse.*

[SUGGEST FOR A ROLE] **4** [S P O **for /as** role] to suggest at a formal meeting that someone should be elected to a particular role: *I propose Tom Ellis for chairman.*

[CELEBRATE] **5** **propose a toast (to sb) | propose sb's health** to ask people to wish sb health, happiness, and success or celebrate their achievement, by raising their glasses and drinking.

The obvious differences are slight. The definitions and examples are mostly unchanged. The overall length is slightly shorter, even though the grammatical apparatus is slightly more elaborate. Small improvements in accuracy and conciseness can be achieved by applying the eight principles outlined above to the existing wording of a verb entry, even to a comparatively 'open' word like *propose*. Much greater improvements are achieved by application of these principles to words at the more 'idiomatic' end of the scale, such as *devour* and *scratch*. There is insufficient room to discuss these here in terms of OALD's entries for these words. Instead, I would ask readers to go straight to the pattern dictionary entries in section 6 of this paper, where a more radical mapping of meaning onto use is proposed, and to make their own comparisons. The entries in all current dictionaries, including OALD, are meaning-driven, i.e. they ask the question, "How many senses does each word have, and what is the definition of each sense?" The question addressed by the pattern dictionary (of which a sample is given in section 6) is, "How many patterns does each word participate in, and what is the sense of each pattern?"

Before moving on to that, however, I would like to comment briefly on grammar in some other dictionaries. The grammatical apparatus of the first edition of the *Longman Dictionary of Contemporary English* (LDOCE; 1978) was at least as elaborate as that of Hornby, but even more impenetrable for ordinary learners. By the corpus-based 3rd edition (1995), LDOCE had adopted a much simpler grammatical apparatus, joining the general trend of learners' dictionaries away from explicit grammar patterns. The only technical terms that it uses are Intransitive [I] and Transitive [T]. The third argument of a verb is described, systematically, as [+ adv/prep], thus:

**amble** *v.* [I always + adv/prep] to walk in a slow relaxed way: [+ **along/across etc**] *the old man came out and ambled over for a chat.*

The front matter of LDOCE3 comments: “You cannot simply say ‘he ambled’ without adding something like ‘along’ or ‘towards me’.”

The *Cambridge International Dictionary of English* (CIDE, 1995), which changed its name to *Cambridge Advanced Learner’s Dictionary* (CALD) for the second edition (2005)<sup>7</sup>, adopts a similar apparatus to that of OALD and LDOCE: minimal and economical, but sufficient.

Surprisingly, the grammatical apparatus of the *Macmillan English Dictionary for Advanced Learners* (MEDAL, 2002) is Spartan to the point of being misleading. This seems to be a deliberate policy, since the principals involved in creating MEDAL had all worked on other learners’ dictionaries, which have more sophisticated, though perhaps minimal, grammar patterns. Presumably, it was decided as a matter of policy that MEDAL should focus on meanings, examples, and collocations, not on grammatical abstractions.

Let me illustrate this with the MEDAL entry for *amble*. Like traditional dictionaries, MEDAL distinguishes transitive and intransitive subcategories of verb senses but, unlike LDOCE, OALD, and CIDE, it generally neglects or misstates the third argument, if there is one. So *amble* in MEDAL is described simply as *verb [I]*. This implies that sentences like *\*the old man ambled is a well-formed sentence of English*. It is not. You have to say where he ambled to—*along, out of the house, into the pub, or whatever*.

This is only one of several examples that could be mentioned. Cumulatively, they add up to a misleading account of verb grammar. It seems that MEDAL has allowed its desire to keep things simple for the learner to be carried to the point where the policy interferes with accurate reporting of the facts of the language.

Cobuild falls into a similar trap. In the second and subsequent editions, the grammar pattern for *amble* in this dictionary is correctly given as “V adv/prep”. Unfortunately, the Cobuild definer forgot to replicate the adv/prep in the definiendum (the first part of the full-sentence explanation). The Cobuild explanation reads: ‘When you *amble*, you walk slowly and in a relaxed manner’. This gives the same mistaken impression as MEDAL. There is a word missing. Cobuild’s full-sentence explanation should read, “When you *amble* somewhere, you walk there slowly and in a relaxed manner.”

In other cases, for example *put*, MEDAL hints at the obligatory third argument, which in this case is really an adverbial of location, but it does so (or tries to do so) only by mentioning specific prepositions, not the relevant clause role, e.g.

**put sth in/on/through/etc. sth**

**put sth into/over/out/etc. sth**

---

<sup>7</sup> Several dictionary publishers in Britain have got into the habit of changing the name of their dictionaries with new editions, even if there is comparatively little alteration. Conversely, when a successful dictionary is totally rewritten, so that it is, in fact, a completely different book, its former title may be retained—and the book may even be published under the name of a long-dead editor. No doubt these things are done for good marketing reasons, but they add to the already difficult complexities of giving accurate bibliographical details for lexicographical works.

This is inadequate, a) because it is verbose and b) because it fails to get the right level of generalization. To focus on specific prepositions is irrelevant, obscuring the equally important fact that the adverbial argument for this verb is often realized by other prepositional phrases and indeed as a single word, e.g. *put it here*, *He put the bin outside*.

However, it must be said that the MEDAL entry is not as inadequate in this respect as American dictionaries of English, for example Merriam Webster's Collegiate (MW), which focuses obsessively, repetitively, and often inaccurately on the transitive/intransitive distinction, while saying nothing at all about the third argument, seemingly being unaware of it. MW implies, for example, that *\*I put the cup* is a well-formed sentence of English. Of course, it is not. An adverbial of location is obligatory—you must say *where* you put it.

The root of this lexicographical problem, like many others in the grammatical apparatus of pre-corpus dictionaries, goes back 1,500 years. Latin grammarians such as Aelius Donatus and Priscian divided verb uses into transitive and intransitive, but they did not recognize adverbials as an essential part of clause structure. English grammarians of the 18th century, presumably under the impression that English is really Latin in disguise, did not recognize them either, and English dictionaries in the 19th and 20th centuries followed the 18th-century grammarians in this and other respects. Some current dictionaries have made no attempt to update their grammatical apparatus or to offer an adequate description of the syntagmatic patterns of word behaviour. Merriam Webster's Collegiate is in this tradition, not only failing to identify the third argument of verbs but also postulating nonexistent intransitive variants of transitive senses, for example:

**put**, *vi.* **1** to start in motion; GO, esp: leave in a hurry.

It is hard to know what to make of this. It seems to imply that *\*John put* and/or *\*the train put* are well-formed sentences of English, meaning “John (or the train) went, or left in a hurry”. But in fact, they mean no such thing: they are both meaningless and ungrammatical. And this definition cannot be an attempt to cover the nautical expression *put to sea*, for that is dealt with in a second sense:

**2** *of a ship*: to take a specified course: <put down the river>.

As with so many of Merriam Webster's minor definitions, in the absence of supporting evidence we must resign ourselves to a state of unresolved bafflement. Similar problems afflict the recording of other grammatical features in all American English dictionaries, for example phrasal verbs and determiners. They do not exist in Latin, so their existence is not explicitly recognized in American dictionaries. This is a shocking state of affairs. The corpus revolution and the grammatical analyses of Quirk and other empirically minded grammarians, which have led to so many improvements in British monolingual dictionaries, have up to now been passed by in American lexicography, suffering as it does under the stranglehold of a market leader that has made little or no investment in serious lexicographical research or innovation for over 40 years.

Let us return to our main theme, namely patterns in EFL dictionaries. It is a pleasure to report that, even though MEDAL does not account for adverbials correctly, it does a good job on phrasal verbs and determiners. MEDAL had the great advantage that the compilers were able to use a state-of-the-art tool for corpus analysis, the Sketch Engine (Kilgarriff et al, 2004), to help them select significant collocations and write definitions reflecting these. The dictionary is peppered with explicit reports on common collocations, e.g.

Words frequently used with *propose*:

*nouns*: change, idea, plan, reform, scheme, solution, theory

If you believe that learners of a language build their own competence analogically on the basis of salient examples, these lists of collocates must be of great benefit. We should bear in mind, however, Hornby's scepticism about the reliability of analogy as a learning tool. The debate about the relative merits of rule-based approaches and analogical approaches to language learning will no doubt continue to run and run for many decades to come.

Here is MEDAL's entry for *propose*:

**propose**

1. [T] *formal* to suggest a plan, idea, or action: *Einstein proposed his theory of general relativity in 1915.* ◇ *I propose going to an early film and having dinner afterwards.* ◇ + **that** *She proposed that we see a marriage guidance counsellor.*
2. [T] to make a formal suggestion in a meeting for people to think about and vote on:  
 ◇ **propose sb for sth** *I propose Sue Wilson for treasurer.* ◇ **propose doing sth** *France has proposed creating a rapid-reaction force to deal with the crisis.*  
**2a. propose a motion** to formally suggest an idea or plan at a meeting.
3. [I/T] to ask someone to get married to you: +**to** *He proposed to her in August.* ◇ **propose marriage** *He proposed marriage, but she refused.*
4. [T] If you propose to do something, you intend or plan to do it: *I propose to tell them the absolute truth.*

It can readily be seen that most of the information that is in OALD6 and in LDOCE is presented here in a similar order and in similar wording, though formatted slightly differently. At sense 3, MEDAL's explicit mention of *propose marriage* is rather more helpful than OALD's formulation, "*~ (sth) (to sb)*". It is arguable, however, that the grammatical label at MEDAL's sense 3 should simply be [I] and the lexically specific transitive alternation 'propose marriage' should be ignored, on the grounds that it is rare, used only for clarification, and covered by sense 1 anyway.

The notion that the infinitive in sense 4 represents a transitive [T] is debatable. It is arguably more helpful to learners to classify the to-infinitive as a clausal argument, and to reserve [T] for noun phrases. However, MEDAL is not alone in taking this view of infinitives: Francis et al. (1996), for example, takes a similar line.

The fourth dictionary I wish to mention in this context is Cobuild. Cobuild has always been a corpus-driven dictionary, so it was spared the expense of having to revise all of its entries in the light of corpus evidence. However, it has made up for this by adopting radically different policies with regard to grammatical description in different editions, and replacing all its examples from new corpus data anyway. The first edition (1987) offered a SPOCA-based description of the clause structure associated with each meaning of each verb. The grammatical apparatus of this first edition received mixed reviews. Admittedly, it was often cumbersome and hard to follow, and occasionally got things wrong or ventured into controversial territory. These may be among the reasons why Cobuild2 adopted a more minimalist, streamlined approach to grammatical description. The grammatical descriptions were moved to sit alongside examples rather than explanations, which yields a great improvement in clarity. However, in addition, the SPOCA-based terminology was abandoned, and the apparatus for grammatical description was reduced to a word-class based system similar to those adopted by later editions of LDOCE and OALD. I cannot help feeling that this move, shared by all EFL dictionaries, has been a case of throwing out the baby with the bathwater. You will see why, I hope, in the discussion of pattern grammar in section 4 below. Simplicity and clarity are great virtues, but not if they are bought at the expense of descriptive adequacy.

### **propose, proposes, proposing, proposed (COBUILD3)**

1. If you propose something such as a plan or idea, you suggest it for people to think about and decide upon:
 

<i>Britain is about to propose changes to European Community institutions.</i>	V n
<i>It was George who first proposed that we dry clothes in that locker.</i>	V that
2. If you propose to do something, you intend to do it.
 

<i>It's far from clear what action the government proposes to take</i>	V to-inf
<i>And where do you propose building such a huge thing?</i>	V -ing
3. If you propose a theory or an explanation, you state that it is possibly or probably true, because it fits in with the evidence that you have considered.
 

<i>This highlights a problem faced by people proposing theories of ball lightning.</i>	V n
<i>Newton proposed that heavenly and terrestrial motion could be unified with the idea of gravity.</i>	V that

4. If you propose a motion for debate, or a candidate for election, you begin the debate or the election procedure by formally stating your support for that motion or that candidate.
- A delegate from Siberia proposed a resolution that he stand down as party chairman.* V that
- I asked Robin Balfour and Derek Haig to propose and second me.* V n
5. If you propose a toast to someone or something, you ask people to drink a toast to them.
- Usually the bride's father proposes a toast to the health of the bride and groom.* V n
6. If you propose to someone, or propose marriage to them, you ask them to marry you.
- He had proposed to Isabel the day after taking his seat in Parliament.* V to n

A unique feature of Cobuild is that it systematically attempts to capture informally the collocational preferences of each sense of each word, by means of ‘full-sentence definitions’, of which the first part is usually the definiendum (the phrase or pattern that is to be defined), encoded within the definition. Cobuild is also “corpus-driven”. Patterns are discovered through corpus analysis. It is, therefore, disappointing to have to note that, in terms of the distinction being made in this paper, the entry structure of Cobuild is meaning-driven rather than pattern-driven. *Proposing an idea* and *proposing a theory*, for example, are treated as separate senses, just as they are in other dictionaries. If the compilers had focused on patterns rather than senses, this dubious semantic distinction might have been treated as a single pattern. Like all other existing major dictionaries, Cobuild’s starting point is a list of senses for each word, not a list of the patterns in which the word normally participates. It was also criticized for a tendency to verbosity. Critics have associated this tendency with the “full-sentence definitions”. In my opinion, the criticism of verbosity was to some extent justified and indeed was addressed in the second edition. However, associating this with full-sentence definitions misses the point. Cobuild is the only serious attempt by any dictionary to systematically identify collocates by semantic type (as opposed to word class), in the definiendum. The impression of verbosity of the definitions results from two factors: firstly, a tendency not to know when to stop, as in the original definition 7 of *proportion* (below) and secondly the frequent attempts to deal with more than one pattern at the same time, as in sense 4 of *propose* (above).

In the first edition of Cobuild (1987), definition 7 of *proportion* read as follows:

If you say that something is big or small **in proportion to** something else, you mean that it is big or small when you compare it with the other thing or measure it against the other thing.

This is undeniably verbose. In the 2001 edition, it was reduced to:

If something is small or large **in proportion to** something else, it is small or large when compared with that thing.

This is a full-sentence definition, but not especially verbose.

MEDAL defines *proportion* (sense 2) as: “the correct, most useful, or most attractive relationship between two things”, and offers the phrase *in proportion to* with an example (“his head is large in proportion to his small frame”) but no definition. An undefined example may be the best strategy for such a phrase.

It is time to move on, but before we leave the question of whether any existing dictionaries are pattern-driven, there is one more dictionary to consider. It is not a learner’s dictionary but a dictionary aimed at native speakers. Somewhat surprisingly, it comes closest to showing how sense distinctions can be made on the basis of patterns. The *New Oxford Dictionary of English* (1998)<sup>8</sup> is (so far) the only dictionary of English aimed at native speakers that both takes corpus evidence seriously and incorporates grammatical descriptions in the Hornby tradition. The entry for *propose* is as follows:

### **propose**

► **verb 1.** [with obj.] put forward (an idea or plan) for consideration and discussion by others: *he proposed a nine-point peace plan* | [with clause] *I proposed that the government should retain a 51 per cent stake in the company.*

■ nominate (someone) for an elected office or as a member of a society: *Roy Thomson was proposed as chairman.*

■ put forward (a motion) to a legislature or committee: *the government put its slim majority to the test by proposing a vote of confidence.*

■ [with infinitive] intend to do something: *he proposed to attend the meeting.*

**2.** [no obj.] make an offer of marriage to someone: *I have already proposed to Sarah* | [with obj.] *one girl proposed marriage to him on the spot.*

In this entry, there is a clear attempt to associate sense distinctions with pattern distinctions, using SPOCA as a basis. Keen-eyed readers will no doubt notice that there is no mention of the expression *propose a toast*.

So far, I have discussed grammar patterns in relation to lexical definitions. But notice that the grammatical error mentioned by Hornby, *\*I proposed him to come*, is not a simple error of structural pattern. The structural pattern “S P O to/INF” (or, if you prefer, “V n to/INF”) is perfectly correct for *propose* in some contexts, e.g. *the Council proposed a plan to widen the road*. The error lies in the selection of a word of the wrong semantic type—the personal pronoun *him*, which has Semantic Type [[Human]] rather than [[Plan]]—in the object slot.

<sup>8</sup> Now marketed in a revised edition as the *Oxford Dictionary of English* (not to be confused with Oxford’s great historical work in lexicography, the 20-volume *Oxford English Dictionary*).

#### 4. Pattern grammar vs. pattern dictionary

There are two possible approaches to using a corpus to identify patterns in text: pattern grammar and pattern dictionary. Both have their merits; both have their shortcomings. The *Pattern Grammar* of Hunston and Francis (2000; H&F) is based on the grammatical apparatus of the second edition of the Cobuild dictionary. It is founded on corpus analysis (i.e. on real texts) and seeks empirically valid generalizations. The following remark (p. 83) is highly relevant:

One of the most important observations in a corpus-driven description of English is that patterns and meanings are connected.

On pages 199-207, H&F discuss grammatical patterns in a short text, reproduced below, which they refer to as “the Joseph Byers text”. In this section of my paper, I shall use this text and the H&F discussion of it to illustrate some of the differences between a pattern dictionary and a pattern grammar and to show how the two approaches are complementary. Here is the text:

Private Joseph Byers was the first Kitchener volunteer to be executed. He was 17 and under age when he enlisted in the 1st Royal Scots Fusiliers in November 1914, and was sent to France with two weeks training. By January 1915, his inexperience and the horrors he witnessed caused him to go absent without leave with another private, Andrew Evans. Byers pleaded guilty, believing that his candour would save him from the death sentence. Despite being under age, he was given no representation at his trial, and he and Evans faced a firing squad at Loche on February 6.

According to rumours, one of them did not die until the third volley, leading to speculation that the firing squad had fired wide to avoid killing the youth.

Table 1 (below) compares the verb patterns identified by H&F in this text with the relevant pattern of each verb in the Pattern Dictionary of the Corpus Pattern Analysis project (CPA, in progress). H&F limit themselves to expressing patterns all at more or less the same level of generalization, almost exclusively in terms of word classes (parts of speech), with the exception of certain prepositions. CPA, by contrast, devotes a great deal of attention to selecting the appropriate level of generalization to capture the meaning of the lexical pattern and to contrast it with other meanings activated by other patterns for the same verb. This necessitates a much richer grammatical apparatus, including identifying, among other things, the semantic type of the subject and, for each clause role, statistically significant collocates grouped by semantic type. Semantic types are identified in double square brackets and refer to a shallow ontology (see section 8 of this paper).

Verb	Pattern Grammar	Pattern Dictionary	Comments
execute	V n	[[Human 1]] execute [[Human 2]]	Semantic types distinguish this sense from others of the same verb, e.g. ‘execute an order’.
enlist	V in n	[[Human]] enlist [NO OBJ] {in [[Human Group = {Military}]}	CPA marks intransitive patterns explicitly. This pattern contrasts with patterns such as “[[Human]] enlist [[Assistance]]”.
send	<i>be</i> V-ed to n (passive of V n to n)	[[Human 1]] send [[Human 2]] [A[Direction]]	
witness	V n	[[Human]] witness [[Event]]	
cause	V n to-inf	[[Anything 1]] cause [[Anything 2]] {to/INF [V]}	In this pattern, semantics add nothing to the basic word-classes.
go	V adj	[[Human]] go [NO OBJ] {absent   AWOL}	Light verb (“delexical verb” in Sinclair’s terminology), with a Subject Complement. This small lexical set activates a particular meaning of <i>go</i> , contrasting with several other patterns of <i>go</i> having a Subject Complement, e.g. <i>go</i> { <i>mad</i>   <i>bananas</i> }.
plead	V adj	[[Human]] plead [NO OBJ] {guilty   {not guilty}}	The adj. in this pattern is a Subject Complement, populated by a lexical set of just two possible items, {guilty} and {not guilty}.
believe	V that	V that [[Human]] believe [NO OBJ] {(that) [CLAUSE]}	
save	V n <i>from</i> n	[[Anything]] save [[Entity]] {from [[Event = Bad]]}	Discussed in Church and Hanks (1990)—a paper not mentioned in Hunston and Francis’s bibliography.
give	be v-ed n (passive of V n n)	[[Human   Event]] give [[Entity 1 = Recipient]] [[Entity 2 = Benefit]]	See discussion of <i>lose</i> below.

Verb	Pattern Grammar	Pattern Dictionary	Comments
face	V n	[[Human]] face [[{Event   Possibility}= Bad]]	Contrasts with [[Entity]] face [A[Direction]]
die	V	[[Animate]] die [NO OBJ] ([A])	Even though an Adverbial is not an obligatory part of the structure of <i>die</i> (and indeed <i>die</i> is often cited as a “one-argument verb”), the norm for <i>die</i> is that it normally governs an <u>optional</u> Adverbial.
lead	V to n	[[Anything]] lead ([[Human]] {o [[Belief]])}	Contrasts with patterns such as [[Route]] lead {to [[Location]]}
fire	V adj	[[Human]] fire [NO OBJ] ([A[Direction]])	H&F fail to identify this pattern correctly. See discussion below.
avoid	V -ing	[[Human   Animal]] avoid [ING]	This pattern contrasts with the pattern [[Human]] avoid [[Event]], e.g. <i>he managed to avoid extradition</i> , where the [[Human]] is typically a Patient not an Agent.

Table 1: Comparison of Pattern Grammar with Pattern Dictionary

It should be mentioned here that the book in the Cobuild series on verb patterns—Francis et al. (1996), which was published four years before H&F—is the one that goes farthest in the delicacy of its grammatical apparatus for syntagmatic distinctions. For example, it groups noun arguments of verbs together according to broad semantic classes, where possible. Thus, under the pattern “V n for n” (Francis et al., 1996: p. 370), there is a meaning group “reward and punish”, which associates this sense of *execute* not only with a direct object and a prepositional phrase (V n for n), but also with the semantic value ‘human’ for both subject and object. It also associates this sense with a third argument—an adverbial governed by *for*—which encodes the thing that the person has done to warrant the reward or punishment. Other members of this meaning group are particular senses of *arrest*, *excuse*, *forgive*, *prosecute*, *punish*, *reward*, *pay back*, *sue*, and *thank* (e.g. *He told officers he wanted to pay them back for locking him up*).

This kind of classification would be sufficient to distinguish *executing a person for murder* from *executing a plan or order*. Unfortunately, however, Francis et al. do not specifically mention this second sense of *execute* under the relevant pattern, “V n”. Their book is a grammar, not a dictionary, so only the most frequent examples of each pattern are given. *Execute a plan* was not common enough to be selected

as a meaning group under “V n”. Francis et al. (1996) is now out of print. It was a pioneering effort in corpus-based grammar and should be revived, as it sketches out some important principles for lexical analysis, which deserve closer study.

Let us now consider the verb *fire* in the fragment ‘speculation *that the firing squad had fired wide*’. I shall go into some detail on this example. The meaning is clear, but how is it constructed? Does it represent a realization of a pattern, or is it anomalous? H&F say:

... we use [the word ‘pattern’] to indicate a sequence of elements that occur with a particular lexical item in this text, whether or not such a sequence is typical. For example, we show the verb *fire* ... with the pattern **V adj**, even though that pattern is productive, is not particularly frequent with this word, and does not distinguish this verb from others.

It seems odd to claim that *any* observed sequence of element can count as a “pattern”. In contrast to what H&F say here, the Pattern Dictionary classifies as patterns only those syntagmatic strings that can be shown, by analysis of corpus evidence, to be typical—i.e. conventional, recurrent chunks of meaningful linguistic behaviour. Classifying just any sequence of elements as a pattern, no matter how idiosyncratic it may be, would seem to defeat the purpose of pattern analysis, opening the floodgates to any observed sequence of elements, no matter how rare or bizarre. In Hanks (forthcoming) a fundamental distinction is made between normal patterns of word use and abnormal uses which deliberately exploit the normal patterns. The latter class includes not only creative metaphors, but also elliptical and anomalous arguments. An example of an anomalous argument of the verb *fire* is ‘stinking spray’ in 1, which exploits the pattern element [[Projectile]], which is populated canonically by words such as bullet, round, shell, rocket, missile, flare.

1. Anyone who has encountered a skunk will know that before it **fires** its stinking spray it issues clear warnings of its intentions.

Be all that as it may, it seems to me that in this particular case, ‘the firing squad fired wide’, there is a pattern, but H&F have failed to identify it correctly. This is because they do not acknowledge clause roles. The pattern in fact consists of a syntagmatic structure with semantic values, expressed as:

[[Human]] fire [NO OBJ] [A[Direction]]

The first hurdle for a lexical analyst in constructing this pattern and applying it to the verb *fire* is to recognize that there is an intransitive verb pattern and that this intransitive pattern is semantically linked to a transitive pattern, “[Human] fire [[Artifact = Firearm]]”. The second hurdle is to recognize that “[Human] fire [NO OBJ] ([A[Direction]])” is a pattern in which a firearm is implied by coercion, even though it is not mentioned explicitly. We can now explain the word *wide*. Contrary to what H&F say, this is in fact not an adjective at all, but an adverbial—a one-word lexical realization of [A[Direction]], a realization of a kind found with several other verbs, for example *aim wide*, *drop short*, *go home*. It answers the question, “Where

did they aim?” or “What did they aim at?” It belongs in the same clause role as *fired over their heads* and *fired into the crowd*.

This analysis goes beyond simple word classes. It introduces contrasts based on the semantic values of collocates, not just syntagmatic structures. In pursuance of this goal, let us ask the sort of question that is asked by the Berkeley FrameNet project, namely: what are the frame elements involved in the semantic frame of people using firearms? We can compile a list like this:

Agent – the person firing the gun

Instrument – the gun or other firearm used

Projectile – the bullet or shell that is fired from the firearm

Target – the thing aimed at or hit

In corpus analysis, frame elements like these are mapped onto idiomatic uses of the lexical items involved. The direct object of the verb *fire* (when the verb is transitive) can be either the Instrument (*fired a gun*) or the Projectile (*fired a shell*).

The difference between a pattern grammar and a pattern dictionary is that a pattern grammar seeks generalizations that affect very large numbers of lexical items, whereas a pattern dictionary looks at each lexical item individually and asks how many patterns it participates in—and what they mean. A pattern dictionary uses patterns to distinguish different meanings of a verb. To do this, it must introduce into the apparatus more delicate structural levels than mere word classes. The theoretical foundations for doing this can be traced back to Halliday (1966) and Sinclair (1966). It seems obvious enough that *firing a gun* activates a different meaning of the verb *fire* from that activated by *firing a person*, even though both these phrases have the structural pattern “V n”. The direct objects must be distinguished according to their semantic types: [[Firearm]] and [[Human]] respectively. Next, discovery procedures are needed to predict whether a lexical item that occurs as the direct object of *fire* is more likely to be a [[Firearm]] or a [[Human]]. To do this effectively, semantic values must be assigned to the arguments of patterns. These semantic values are encoded in a shallow ontology, which I will discuss in the next section.

## 6. Introducing semantic values of arguments into patterns

Consider for a moment the verb *enlist*. It has two senses. The H&F pattern grammar rightly shows that ‘enlist in the army’ (grammar pattern: V *in* n) and “enlist someone’s help” (grammar pattern: V n) have different meanings. Here, the pattern dictionary and the pattern grammar agree. But this is only the tip of the iceberg of verb meaning distinctions. Many competing meanings of verbs have precisely identical patterns in terms of the limited apparatus of grammatical analysis that H&F use, as we have seen. Meaning distinctions very often depend on a distinctive semantic type of one or more of the arguments. For example, *firing bullet from a*

*gun* and *firing a person from a job* can both be described as “V n from n”, using the terminology of pattern grammar. To get the meaning distinction, we need to invoke a more delicate analytic level than mere word classes. The two different meanings of *fire* are activated by differences of semantic type in the direct object slot, namely [[Human]] and [[Projectile]]. This distinction is confirmed by differences of semantic type in the prepositional object slot, namely [[Firearm]] and [[Activity]]. The majority of semantic distinctions for polysemous verbs are of this kind, not the ‘enlist’ kind.

Sometimes, it is the distinction in semantic type of a prepositional object that makes all the difference. For example, consider the verb *sail*. One common use of this verb is to *sail through* something. Here we have a verb + preposition. Is this sufficient evidence to decide the meaning of the verb *sail*? No! It is necessary also to know the semantic type of the prepositional object. Consider the following three examples:

1. In 1577 he set out in the Pelican (afterwards renamed the Golden Hind) for the river Plate, **sailed through** the Straits of Magellan, plundered Valparaíso, rounded the Cape of Good Hope, and completed the circumnavigation of the world .
2. Jeremy Irons, who **sails through** the role with charm and panache.
3. I’ve even heard 12-year-olds **sail through** this work [Samuel Barber’s violin concerto]

The meaning depends on the semantic type of the prepositional object: [[Location]] vs. [[Activity]]. It may be objected that acting roles and violin concertos are not activities. This is true, but irrelevant. It overlooks the fact that what is meant in 2 is the *acting* of the role and in 3 the *playing\_of* Barber’s violin concerto. Playing a concerto is, of course, an activity. These are examples of the kind of semantic coercion, a notion introduced by Pustejovsky (1995). Nouns like *role* and *work* are coerced by the verb+preposition combination *sail through* into having the semantic type of the activity most normally associated with them: acting and playing. This is how the prepositional objects of 2 and 3 activate the ‘accomplish with ease’ sense of *sail through*, while the prepositional object in 1 activates the sense ‘pass through in a boat’. In example 1, “through the Straits of Magellan” is just one of many adverbial of direction governed by the verb *sail* in this sense, while 2 and 3 are much more idiomatic constructions.

Introducing the semantic types of lexical items as an analytic level is necessary, but it unleashes a veritable hailstorm of problems for the lexical analyst. The only dictionary which has even attempted to capture relevant meaning-determining collocations at this level is Cobuild. Cobuild is defective in many ways, but it at least made a start on addressing the question of how word meaning is related to word use. Without corpus evidence and statistical measures of collocational salience, the question cannot be addressed seriously at all.

Patterns are useless if they do not have predictive power. Therefore, for each lexical item in each pattern of each verb, we need to ask. “How likely is it that we will see this word again, as a collocate of our target word, in a comparable expanse of text?” Answers can be computed statistically on the basis of large samples. *CPA (Corpus Pattern Analysis)* regards a pattern as a relationship between sets of collocates. A single word cannot be a pattern.

Verb patterns consist of two or more words in a syntagmatic structure. There is also a paradigmatic element in a verb pattern: the arguments consist of lexical sets of nouns or other words. Typically, these lexical sets are sets of synonyms. Adjective patterns are also syntagmatically structured: the adjective is either a modifier of a particular set of nouns or is related to a set of nouns and structures by a linking verb such as *be* or *seem*. The semantic analysis of adjectives is much like that of verbs in this respect.

Noun patterns, however, do not necessarily have a syntagmatic structure. Significant collocates can be in an unstructured relationship with one another and still function in the same way as structural patterns in assigning probabilities to the selection of a relevant meaning of a target word. To take a simple example, the noun *doctor* has at least two senses: 1) medical practitioner, and 2) bearer of an advanced academic degree. The first sense is much commoner, and is typically distinguished by collocation with any of a very large number of words such as *patient, dentist, surgeon, nurse, treat, symptom, or hospital*. If these words are found anywhere close to the target word (*doctor*), it is a fair bet that the medical sense is the one that should be selected. On the other hand, if *doctor* occurs near words such as *degree, philosophy, divinity, or letters*, the rarer academic sense is more likely to be the correct one.

I hasten to add that in the first sentence of the preceding paragraph emphasis must be placed on ‘not necessarily’. It is undeniable that many nouns, especially nouns that are derived from verbs, do have a syntagmatic structure. But unstructured collocation is a phenomenon more associated with nouns than with verbs.

Rather than prolonging the theoretical discussion, I will conclude this section by quoting some examples of entries from the Pattern Dictionary. I will not discuss the points raised by these entries in any great detail, as a whole workshop would be needed to do that properly. A fuller discussion of the aims of the Pattern Dictionary and a contrastive study with FrameNet and other work will be found in Hanks and Pustejovsky (2005).

Pattern elements vary greatly in scope: a pattern element may be any of the following:

- a) a whole phrase (e.g. an adverbial of direction)
- b) a cluster of nouns sharing the same semantic type or other attribute (e.g. [[Human]]), or
- c) an individual word (typically, individual words are pattern elements of idioms).

The patterns for each verb aim at being mutually exclusive: that is, if the nouns, adjectives and other words that realize each argument of a verb in an unseen clause are assigned to the right semantic type—i.e. the right place in the project’s shallow ontology—then the meaning of the clause as a whole can be identified with reasonable confidence. Meanings are expressed as **implicatures**.<sup>9</sup> Each implicature is ‘anchored’ to the corresponding pattern by replication of pattern elements in both pattern and implicature. Not all pattern elements are replicated in the implicature, however. Idioms, in particular, are very weakly anchored. The converse is also true: occasionally, a pattern element is found only in an implicature and not in the pattern itself. This happens when an argument is strongly implied by a verb even though it is not explicitly present in the clause structure.<sup>10</sup>

The first example is the verb *amble*, discussed above, which has only one sense and one pattern. There are no surprises here. The purpose of showing it is merely to begin to familiarize the reader with the metalanguage of the Pattern Dictionary.

### **amble**

1. PATTERN: [[Human | Animal]] amble [NO OBJ] [A[Direction]]  
PRIMARY IMPLICATURE: [[Human | Animal]] walks slowly and in a relaxed manner in a certain [[Direction]]  
COMMENT: [A[Direction]] is almost invariable present in the syntagmatics, although semantically it is unimportant, as the focus of this verb is on manner of motion, not on the direction of movement.  
EXAMPLE: Two sheep and a goat ambled up over the roof and grazed on its turf.

Notice that the primary implicature is anchored to the pattern by repetition in both places of as many clause roles as possible. In this way, a link is established between meaning and use.

The next example is the entry for the verb *devour*. This, too, is fairly straightforward. It has four lexico-semantic patterns, all of them realizations of the “V n” syntactic structure. Note that the basic sense is manner of eating, not just eating, and this has give rise to two other conventional patterns: a person devouring a book and one institution devouring another.

---

<sup>9</sup> The primary implicature is closest to a dictionary definition. But a verb pattern is a hook onto which any number of secondary implicatures can be hung.

<sup>10</sup> For example, the phrasal verb *bandy words around* implies an exchange of words between two people, acting alternately as audience and utterer, even though normally only one of them is explicitly realized in any given sentence.

### devour

1. PATTERN: [[Human 1 | Animal 1]] devour [[{Animal 2 = Food} | {Physical Object = Food}]]  
 PRIMARY IMPLICATURE: [[Human 1 | Animal 1]] hungrily eats [[Animal 2 = Food | {Physical Object = Food}]]  
 SECONDARY IMPLICATURE: [[Human 1 | Animal 1]] eats all of [[{Animal 2 | Physical Object} = Food]], so that nothing is left  
 EXAMPLE: *Prince Khalid Bin Sultan ... is said to have turned pale when Egyptian commandos **devoured** live chickens and rabbits in a show of bravado. | Here a swarm of common starfish are rapidly **devouring** the carcass of a fish.*  
 FREQUENCY: 58%
2. PATTERN: [[Human]] devour [[Document]]  
 PRIMARY IMPLICATURE: [[Human]] eagerly reads [[Document]]  
 EXAMPLE: *The author's explanation of why people **devour** books about the rich is appropriately cynical.*  
 FREQUENCY: 14%
3. PATTERN [[Human | Institution 1 | Abstract 1]] devour [[Institution 2 | Abstract 2]]  
 PRIMARY IMPLICATURE: [[Human | Institution 1 | Abstract 1]] takes over, uses, absorbs, and destroys [[Institution 2 | Abstract 2]]  
 EXAMPLE: *No peaceful international order is possible if larger states can **devour** their smaller neighbours.*  
 FREQUENCY: 24%

My final example is the verb *scratch*. This is more complex. 14 patterns may be distinguished. Some of the distinctions are quite fine-grained, but they are of vital importance in answering the question “Who did what to whom?”. No distinction is made between semantic and pragmatic implicatures, though secondary implicatures often express pragmatics. As far as CPA is concerned, they are all part of the conventional meaning of these expressions.

### scratch

1. PATTERN: [[Human | Physical Object 1]] scratch [[Physical Object 2]]  
 PRIMARY IMPLICATURE: [[Human | Physical Object 1]] marks and/or damages the surface of [[Physical Object 2]]  
 SECONDARY IMPLICATURE: Typically, if subject is [[Human]], [[Human]] does this by dragging a fingernail or other pointed object across the surface of [[Physical Object 2]]  
 EXAMPLE: *I remember my diamond ring **scratching** the table. | 'I'm sorry sir, but I'm afraid I've **scratched** your car a bit!'*  
 FREQUENCY: 19%

2. PATTERN: [[Human]] scratch [[Language | Picture]] {on [[Inanimate = Surface]]}  
 PRIMARY IMPLICATURE: [[Human]] writes or marks [[Language | Picture]] on [[Inanimate = Surface]] using a sharp edge or pointed object  
 EXAMPLES: *A Turkish schoolboy who had **scratched** the word 'Marxism' on his desk.* | *Names of infant Mulverins had recently been **scratched** on the wall.*  
 FREQUENCY: 9%
3. PATTERN: [[Human | Animal]] scratch [[Self | Body Part]]  
 PRIMARY IMPLICATURE: [[Human | Animal]] repeatedly drags one or more of his or her fingernails rapidly across [[Body Part]]  
 SECONDARY IMPLICATURE: typically, [[Human | Animal]] does this in order to relieve itching  
 EXAMPLE: *Without claws it is impossible for any cat to **scratch** itself efficiently.*  
 FREQUENCY: 16%
4. PATTERN: [[Human]] scratch {head}  
 PRIMARY IMPLICATURE: [[Human]] rubs his or her {head} with his or her fingernail(s)  
 SECONDARY IMPLICATURE: often a sign that [[Human]] is puzzled or bewildered  
 EXAMPLE: *He peered down at me and **scratched** his head as he replaced his cap* | *Having just struggled through a copy of the Maastricht Treaty I can only **scratch** my head that anyone would wish to sign it* [METAPHORICAL EXPLOITATION].  
 FREQUENCY: 14%
5. PATTERN: [[Human 1 | Animal 1]] scratch [[Human 2 | Animal 2]]  
 PRIMARY IMPLICATURE: [[Human 1 | Animal 1]] uses the fingernails or claws to inflict injury on [[Human 2 | Animal 2]]  
 EXAMPLES: *Mary was starting to pull her sister's hair violently and **scratch** her face in anger.*  
 FREQUENCY: 9%
6. PATTERN: [[Inanimate]] scratch [[Human | Animal]]  
 PRIMARY IMPLICATURE: [[Inanimate]] accidentally inflicts a superficial wound on [[Human | Animal]]  
 EXAMPLE: *A nice old Burmese woman brought us limes—her old arms **scratched** by the thorns.*  
 FREQUENCY: 2%

7. PATTERN: [[Bird = Poultry]] scratch [NO OBJ] (around)  
 PRIMARY IMPLICATURE: [[Bird = Poultry]] drags its claws over the surface of the ground in quick, repeated movements  
 SECONDARY IMPLICATURE: typically, [[Bird = Poultry]] does this as part of searching for seeds or other food.  
 EXAMPLE: *A typical garden would contain fruit and vegetables, a few chickens to **scratch** around*  
 FREQUENCY: 3%
8. PATTERN: [[Human]] scratch [NO OBJ] {around | about} {for [[Entity = Benefit]]}  
 PRIMARY IMPLICATURE: [[Human]] tries to obtain [[Entity = Benefit]] in difficult circumstances  
 COMMENT: Phrasal verb.  
 EXAMPLE: *Worrying his head off, **scratching** about for the rent*  
 FREQUENCY: 4%
9. PATTERN: [[Human]] scratch {living}  
 PRIMARY IMPLICATURE: [[Human]] earns a very poor {living}  
 COMMENT: Idiom.  
 EXAMPLE: *destitute farmers trying to **scratch** a living from exhausted land.*  
 FREQUENCY: 6%
10. PATTERN: [[Human 1]] scratch {[[Human 2]]'s {back}}  
 PRIMARY IMPLICATURE: [[Human 1]] helps [[Human 2]] in some way  
 SECONDARY IMPLICATURE: usually as part of a reciprocal helping arrangement  
 COMMENT: Idiom.  
 EXAMPLE: *Here the guiding motto was: you **scratch** my back, and I'll scratch yours—a process to which Malinowski usually referred in more dignified language as 'reciprocity' or 'give and take'.*  
 FREQUENCY: 1%
11. PATTERN: [[Human | Institution]] scratch {surface (of [[Abstract = Topic]])}  
 PRIMARY IMPLICATURE: [[Human | Institution]] pays only very superficial attention to [[Abstract = Topic]]  
 COMMENT: Idiom.  
 EXAMPLE: *As a means of helping Africa's debt burden, ... it barely **scratches** the surface of the problem.*  
 FREQUENCY: 11%

12. PATTERN: [[Human 1]] scratch [[Entity]]  
 PRIMARY IMPLICATURE: [[Human 1]] looks below the obvious superficial appearance of something ...  
 SECONDARY IMPLICATURE: ... and finds that the reality is very different from the appearance.  
 COMMENT: Imperative. Idiom.  
 EXAMPLE: *Scratch any of us and you will find a small child.*  
 FREQUENCY: 2%
13. PATTERN: [[Human | Physical Object 1 | Process]] scratch [[Physical Object 2 | Stuff]] {away | off}  
 PRIMARY IMPLICATURE: [[Human | Physical Object 1 | Process]] removes [[Physical Object 2 | Stuff]] from a surface by scratching it  
 COMMENT: Phrasal verb.  
 EXAMPLE: *First he scratched away the plaster, then he tried to pull out the bricks.*  
 FREQUENCY: 2%
14. PATTERN: [Human]] scratch [[Language | Picture]] {out}  
 PRIMARY IMPLICATURE: [[Human]] deletes or removes [[Language | Picture]] from a document or picture  
 COMMENT: Phrasal verb.  
 EXAMPLE: *Some artists ... use 'body colour' occasionally, especially solid white to give that additional accent such as highlights and sparkles of light on water which sometimes give the same results as scratching out.*  
 FREQUENCY: 1%

## 7. An ontology of shimmering lexical sets

I am arguing here that, in order to understand how meaning in language works, it is necessary to start by analysing verbs in context, using a large corpus. The first step is to distinguish the normal, conventional uses of each verb from abnormal, unusual uses. Abnormal uses are set aside for later analysis. To find the conventional uses of verbs, we first identify the different structural patterns—relationships among clause roles—of the kind described by Hornby and his successors. Then each structural pattern is subdivided according to the semantic types of the words in the clause roles, insofar as these activate different meanings of the verb. This work is greatly facilitated by selection of statistically significant or 'salient' collocates in each clause role. There are now tools (in particular, the Sketch Engine) that make it possible to instantly identify salient collocates in different clause roles and other syntagmatic relationships for any content word in any corpus of any language. Grouping these significant collocates into clusters implies that the nouns (at least) must be grouped into an ontology according to their semantic type. How is this to be done?

An ontology or thesaurus is called for, in which words are organized in a semantic hierarchy: the sort of apparatus first implemented by Wilkins (1668) and subsequently by Roget (1853) and Miller and others (1995). Attempts to adopt existing ontologies for CPA proved unsatisfactory, so currently considerable effort is being put into building a shallow ontology that reflects how nouns are actually used in relation to verbs. A prototype of this ontology is outlined in Pustejovsky et al. (2004). The top type is called [[Anything]]. When this is used in a pattern, it means that absolutely any noun, without reference to its semantic classification, can be used in that particular clause role.

The top levels of the CPA Ontology, in a somewhat simplified and schematized version, look like this:

```

Anything
  Eventuality
    Event
    State
  Entity
    Physical Object
      Inanimate
        Artifact
      Animate
        Human
        Animal
        Plant
    Abstract
  
```

It will be seen that:

The top type (Anything) is divided into Entities and Eventualities.

Eventualities are divided into Events and States.

Entities are divided into Physical Objects, Abstracts.

Physical Objects are divided into Animates and Inanimates.

... and so on. There are many more subdivisions and interrelationships.

The terms used in the ontology are not to be thought of as English words, but rather as addresses which will be populated with words. Interesting questions arise when we come to populate the addresses with actual words.

An ontology is usually considered to be an ordered set of hyponyms, synonyms, and co-hyponym, which are in a fixed relationship to one another because they share certain properties of meaning.

For example, a bird is a living creature or [[Animate]], so this word and its synonyms and hyponyms belong in the ontology somewhere under [[Animate]]. Synonyms of *bird* are very few: in fact, its only true synonym is the rather archaic word *fowl*. Hyponyms of *bird*, on the other hand, are plentiful. They include *sparrow*, *finch*,

*osprey*, *hawk*, *penguin*, and a very large number of other words. Already, we can sense trouble ahead, for whereas *sparrow*, *finch*, *osprey*, *hawk* all activate a particular sense of the verb *fly*, the noun *penguin* does not; it is more associated with the verbs *swim* and *waddle*. On the other hand, when we are analysing the verb *breed*, *penguin* re-joins the set of entities that breed (in the sense ‘have offspring’). Thus, in relation to different verbs, some members of a lexical set drop out, while, when we move on to a different verb, other members come in. In this sense, a lexical set may be said to “shimmer”. Its membership is not constant, but changeable. Nevertheless, lexical sets of nouns, in a hierarchically organized ontology, are necessary to pick out different meanings of verbs. The hierarchical organization is necessary because different verbs take arguments at different levels of generality.

One important variable that must be mentioned here is the tension between the principle of idiomaticity and the principle of openness. Sinclair (1991) identifies a tension between what he calls the *open-choice principle*:

a way of seeing language as the result of a very large number of complex choices. At each point where a unit is complete (a word or a phrase or a clause), a large range of choices opens up and the only restraint is grammaticalness

and the *idiom principle*:

Many choices within language have little or nothing to do with the world outside. ... A language user has available to him or her a large number of semi pre-constructed phrases that constitute single choices.

Consider the verb *abandon*. The vast majority of uses of this verb represent a simple transitive structure, i.e. S V O. The subject is normally [[Human]], but what about the direct object? You can abandon an activity, plan, or project—all words that belong in the [[Event]] hierarchy—or a refrigerator, a car, or a TV set—words which come under [[Physical Object]]. You can also abandon [[Human]]s, e.g. your friends or your wife and children—and you can abandon a [[Location]] such as a hilltop or a defensive position. You can also abandon something that are [[Abstract]] such as a scientific theory or an ideology. So *abandon* seems to be a good example of an open-choice verb. On the other hand, the implicatures of *abandoning one’s wife and children* are quite different from those of abandoning a hypothesis or a fortress. There is a general overall sense (‘go away from and no longer have anything to do with X’), but there are also a number of specific implicatures associated with different types of thing that are abandoned. So having grouped the direct objects according to their semantic types, the lexical analyst still has to decide whether to lump or split the senses of *abandon*, and if splitting, how delicate the splits should be. There is no simple right-or-wrong answer to this question: the decision must be motivated by the degree of delicacy required by the intended user or application.

Not only is there variation in co-hyponyms when an ontology is applied to real texts, but also there is also variation in focus. Quite often, a noun denoting a part or a property is used in alternation with a noun denoting a whole entity. Take the verb *calm* as an example. Typically, you calm an animate entity such as a person or

a horse, although in fact, only a subset of animate entities normally occurs as the direct object of *calm*. You do not, for example, normally talk about calming *insects* or *spiders*. However, the direct object slot is also very often used to focus on relevant properties of an entity. You can, without change of meaning, calm people's fears or anxieties. *Fear* and *anxiety* are not animate entities; they are properties of animate entities. They focus on the relevant property of the person or animal concerned; they do not activate a different sense.

With other verbs, the focus may be on parts of the whole. You can *repair* a house or a car, but you can also, in the same sense, repair the roof or windows of a house or car, or some other part such as the headlights of a car or the brickwork of a house.

Then there are words whose semantics cut across semantic classes, e.g. *pet*. Some but not all mammals are pets; some but not all birds are pets. Pet is a role assigned to individuals, not a semantic class within a scientific classification of the universe—and yet, nevertheless, there is a fairly distinctive class of pets, with distinctive properties.

Clustering of lexical items in verb arguments is an important (though up to now neglected) topic in lexical analysis. It needs to be matched with the traditional semantics of lexical sets, as found in thesauruses and ontologies. But, as we have seen, quite a sophisticated analytical apparatus will be required to group words into relevant sets, and not all decisions can be made by algorithm: some lexicographical judgement will always be called for.

## 8. Applications

There is neither sufficient time nor space here to engage in a full discussion of all the potential applications of the CPA Pattern Dictionary, but a brief sketch may help to set the project in perspective. It is not intended as a dictionary for foreign learners or, indeed, any ordinary every human user. The apparatus of brackets and implicatures, I am told, looks intimidating to ordinary users, although in fact it is really quite simple.

The main purpose of the project is to provide empirically well-founded links between word meaning and word use. To do this, it proceeds via patterns of use, which can be recognized explicitly and measured. It is, therefore, an infrastructure resource with a great many potential applications. These include:

- Computational natural language understanding systems
- Machine translation (associating meaning with patterns in two languages, rather than words)
- Anomaly detection—distinguishing unusual words and expressions from normal phraseology
- Semantic web—processing meaning in unstructured text
- Natural language generation—idiomatic phraseology

- In lexicography: future dictionaries with a much clearer focus on normal phraseology as well as meaning
- Pedagogical applications, including automatic error identification.

On the computational side, Rumshisky (forthcoming) reports on the use of pattern elements to identify automatically the correct meaning of polysemous verbs in free text. She deals with automatic identification of semantically diverse lexical sets that activate the same sense of the predicate. Semantically diverse nouns are grouped into lexical sets on the basis of association with sets of ‘selectionally equivalent’ verbs (i.e. verbs that share selectional preferences in a given argument position).

## 9. Conclusion

It is a truism that context determines meaning, but it is hard to decide what counts as a meaning and what counts as a context. This is an abiding problem for lexicographers, language learners, translators, and computational language processing alike. The problem of identifying context and meaning in unseen text is the theme of this paper.

A. S. Hornby pointed us in the right direction by drawing attention to the highly patterned nature of language in use and by constructing a framework of structural patterns to which different meanings and different idiomatic uses of each verb could be related. He successfully identified the clause structures involved, though these have subsequently been revised and streamlined by his successors. However, Hornby’s patterns took no account of the semantic types of the arguments of verbs. With the resources that were available during his lifetime, he was not able to go much further than analysis of clauses in terms of clause roles and part-of-speech classes, even if he had wanted to. Since then there have been some improvements in clarity, as well as some retrograde steps such as the substitution of analysis in terms of word classes for analysis in terms of clause roles.

In this paper, I have proposed a return to clause roles (rather than word classes) as an essential first step before proceeding to a more sophisticated analysis which systematically relates word meaning to word use. Future lexicography will, I predict, include projects that are pattern-driven rather than meaning-driven. It will include analysis of verb meaning in relation to the semantic types of clause roles, not merely structural patterns of word classes.

To see how uses of a particular verb in different patterns have different meanings, it is necessary to first find the verb, then correlate the grammatical patterns of the verb with its salient collocates, which are grouped together according to different aspects of their semantics. Current priorities for the Pattern Dictionary project include creation of an empirically well-founded ontology and a methodology for representing entities, their properties, and their parts, as meaning-determining collocates in different argument positions.

Identifying statistically significant collocations, grouping them into patterns, and building an ontology are future tasks for lexicographers, corpus analysts, and computational linguists, working together hand in hand.

## Acknowledgements

I would like to thank the Hornby Trust and the organizers of Euralex 2008 for the honour of inviting me to give this lecture. Thanks are also due to Gill Francis and Gilles-Maurice de Schryver for comments on earlier drafts of this paper.

The research was funded in part by the Academy of Sciences of the Czech Republic (project T100300419) and the Czech Ministry of Education (National Research Program II project 2C06009).

## References

### *Dictionaries*

- [CIDE1]. Procter, P. and others (1995). *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press.
- [CIDE2 (CALD)]. Woodford, K. et al. (2005). *Cambridge Advanced Learner's Dictionary* (= 2nd edition of CIDE). Cambridge: Cambridge University Press.
- [COBUILD1]. Sinclair, J. M.; Hanks, P. and others (1987). *Collins Cobuild English Language Dictionary*. London: HarperCollins.
- [COBUILD2, 3]. Sinclair, J. M.; Fox, G.; Francis, G. and others (<sup>2</sup>1995, <sup>3</sup>2001). *Collins Cobuild English Language Dictionary*. London: HarperCollins.
- [COD1]. Fowler, H. W.; Fowler, F. G. (1911). *Concise Oxford Dictionary of Current English*. Oxford: Oxford University Press.
- [COD8]. Allen, R. and others (1990). *Concise Oxford Dictionary of Current English, 8th Edition*. Oxford: Oxford University Press.
- [ISED]. Hornby, A. S.; Gatenby, E. W.; Wakefield, H. (1942). *Idiomatic and Syntactic English Dictionary*. Kaitakusha. Republished in 1948 by Oxford University Press as *A Learner's Dictionary of Current English*. See OALD.
- [LDOCE1]. Procter, P. and others (1978). *Longman Dictionary of Contemporary English*. Harlow: Longman.
- [LDOCE3]. Rundell, M. and others (1995). *Longman Dictionary of Contemporary English, 3rd edition*. Harlow: Longman
- [LDOCE4]. Bullon, S. and others (2003). *Longman Dictionary of Contemporary English, 'new edition'*. Harlow: Longman
- [MEDAL]. Rundell, M. (2002). *Macmillan English Dictionary for Advanced Learners*. London: Macmillan.
- [NODE (ODE)]. Hanks, P.; Pearsall, J. and others (1998). *New Oxford Dictionary of English*. Oxford: Oxford University Press. (2nd edition 2003 published as *Oxford Dictionary of English*.)

- [OALD2]. Hornby, A. S. and others (1962). *Oxford Advanced Learner's Dictionary of Current English, 2nd edition*. Oxford: Oxford University Press. (Second edition of ISED).
- [OALD3]. Hornby, A. S.; Cowie, A. and others (1974). *Oxford Advanced Learner's Dictionary of Current English, 3rd edition*. Oxford: Oxford University Press.
- [OALD4]. Cowie, A. and others (1974). *Oxford Advanced Learner's Dictionary of Current English, 4th edition*. Oxford: Oxford University Press.
- [OALD5]. Crowther, J. and others (1995). *Oxford Advanced Learner's Dictionary of Current English, 5th edition*. Oxford: Oxford University Press.
- [OALD6]. Wehmeier, S. (ed.) and others (2000). *Oxford Advanced Learner's Dictionary of Current English, 6th Edition*. Oxford: Oxford University Press.

### **Ontologies**

- Leibniz, G. W. (1704). *Table of Definitions*. Excerpts, selected and edited by Emily Rutherford. In Hanks, P. (ed.) (2008). *Lexicology: Critical Concepts in Linguistics*. Vol. 1. London: Routledge.
- Miller, G.; Fellbaum, C. and others (1995). *WordNet*. At <http://wordnet.princeton.edu>.
- Roget, P. M. (1852). *Roget's Thesaurus*. Many subsequent editions and imitations.
- Wilkins, J. (1668). *Essay Towards a Real Character and a Philosophical Language, part II: Universal Philosophy*. London: the Royal Society. Excerpts in Hanks, P. (ed.) (2008). *Lexicology: Critical Concepts in Linguistics*. Vol. 1. London: Routledge.

### **Grammars**

- Aelius Donatus. 4th century AD. *Ars grammatica*.
- Dixon, R. M. W. (1991). *A New Approach to English Grammar, on Semantic Principles*. Oxford: Oxford University Press.
- Francis, G.; Hunston, S.; Manning, E. (1996). *Collins Cobuild Grammar Patterns 1: Verbs*. London: HarperCollins.
- Hornby, A. S. (1954). *A Guide to Patterns and Usage in English*. Oxford: Oxford University Press.
- Hunston, S.; Francis, G. (2000). *Pattern Grammar*. Amsterdam: John Benjamins.
- Priscian. c. 500 AD. *Institutiones grammaticae*.
- Quirk, R.; Greenbaum, S.; Leech, G.; Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

### **Other literature**

- Baker, C. F.; Fillmore, C. J.; Cronin, B. (2003). "The Structure of the FrameNet Database". *International Journal of Lexicography* 16 (3). 281-296.
- Halliday, M. A. K. (1966). "Lexis as a linguistic level". In Bazell, C. E.; Catford, J. C.; Halliday, M. A. K.; Robins, R. H. (eds.). *In Memory of J. R. Firth*. London: Longman. 148-162.
- Hanks, P. (forthcoming). *Lexical Analysis: Norms and Exploitations*. MIT Press.

- Hanks, P.; Pustejovsky, J. (2005). "A pattern dictionary for natural language processing". *Revue Française de Linguistique Appliquée* 10 (2). 63-82.
- Miller, G.; Fellbaum, C. (1985-91). "Five papers on WordNet". At: <ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps>.
- Pustejovsky, J. 1995. *The Generative Lexicon*. Chicago: The MIT Press.
- Pustejovsky, J.; Hanks, P.; Rumshisky, A. (2004). "Automated Induction of Sense in Context". In *Proceedings of COLING 2004*, Geneva, Switzerland.
- Rumshisky, A. (forthcoming). "Resolving Polysemy in Verbs: Contextualized Distributional Approach to Argument Semantics". In Lenci, A. (ed.). *From Context to Meaning: Distributional Models of the Lexicon in Linguistics and Cognitive Science*. Special Issue of the *Italian Journal of Linguistics*.
- Sinclair, J. M. (1966). "Beginning the study of lexis". In Bazell, C. E.; Catford, J. C.; Halliday, M. A. K.; Robins, R. H. (eds.). *In Memory of J. R. Firth*. London: Longman. 148-162.
- Sinclair, J. M. (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

#### **Web sites**

- [CPA]. Corpus Pattern Analysis: <http://nlp.fi.muni.cz/projects/cpa/>.
- FrameNet: <http://framenet.icsi.berkeley.edu/>.