Terminology, Phraseology, and Lexicography[1]

Patrick Hanks

Institute of Formal and Applied Linguistics, Charles University in Prague

This paper explores two aspects of word use and word meaning in terms of Sinclair's (1991, 1998) distinction between the open-choice principle *(or* terminological tendency*) and the* idiom principle *(or* phraseological tendency*). Technical terms such as* strobilation *are rare, highly domain-specific, and of little phraseological interest, although the texts in which such word occur do tend to contain interesting clusters of domain-specific terminology. At the other extreme, it is impossible to know the meaning of ordinary common words such as the verb* blow *without knowing the phraseological context in which the word is used.*

Many words have both a terminological tendency and a phraseological tendency. In some cases the two tendencies are in harmony; in other cases there is tension between them. The relationship between these two tendencies is investigated, using examples from the British National Corpus.

1. Introduction

Sinclair (1991) makes a distinction between two aspects of meaning in texts, which he calls the *open-choice principle* and the *idiom principle*. In this paper, I will try to show why it is of great importance, not only for understanding meaning in text, but also for lexicography, to try to take account of both principles even-handedly and assess the balance between them when analyzing the meaning and use of a word. The *open-choice principle* (alternatively called *the terminological tendency*) is:

> 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.

The *idiom principle* (alternatively called the *phraseological tendency*), on the other hand, is described in the following terms:

> 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.

A serious monolingual dictionary must account for both these aspects of word use and word meaning. Up to now, monolingual dictionaries, like term banks, have focused mainly on the terminological tendency - i.e. the meaning of words in relation to the world - and have seriously neglected phraseology. Indeed, until large corpora became available in the 1990s, it was not possible to give an accurate account of the 'semi pre-constructed phrases' that we now know are associated with each word. Those dictionaries which made a serious attempt, on the basis of introspection, to show the phraseology - for example, the heroic *Wörterbuch der deutschen Gegenwartssprache*, compiled in the German Democratic Republic during the Communist era, can now be shown by corpus evidence to have been at best only partly successful. And during the past twenty years, monolingual lexicography has, by and large, tended to neglect the opportunities offered by corpus evidence and to cling to outmoded models of lexicography.

In this paper I explore some implications for lexicography of these two tendencies. In Hanks (2009) I tried to show at least one way in which things might be done differently. Here I discuss certain aspects of the theoretical background and the implications for electronic terminological lexicography on the one hand and electronic phraseological lexicography on the other hand (Hanks and Pustejovsky 2005). Some words are very terminological, others are very phraseological. In some case there is a tension between the two tendencies, in other cases, harmony.

2. The Terminological Extreme

Terminology in its purest form is rare in general language and typically found only in highly specialized texts. An example is the noun *strobilation* and its related verb *strobilate*. These words are not in the British National Corpus or the much larger Oxford English Corpus. They are not in the *Oxford Dictionary of English* (ODE, 2005). They are probably not in the active or passive vocabulary of any user of English who is not a zoologists, or, more specifically, a parasitologist (or a lexicographer). Nevertheless, like many highly specialized technical terms, they are alive and well and living in small clusters of technical domains. A Google search (30 October 2009) yields quite a few citations, for example (from the Cabrillo High School Aquarium project, California, http://www.cabrilloaquarium.org/aqurium-exhibits/moon-jelly-strobilation.html):

Jellies live their lives in two body forms or stages. There is a **polyp** or attached stage which resembles tiny sea anemones (jellies and anemones are actually related) and a **medusa** or free-swimming stage. The trick to getting jellies to go through their life cycle (both stages) is to get the attached polyps to begin to **strobilate**, or **bud off** new **juvenile jellies**. This may sound simple enough, but the polyps are very comfortable just staying as polyps. You actually have to 'trick' them into budding by changing some of their environmental conditions such as temperature, water chemistry and/or nutritional factors.
Head Curator Lisa Scott (class of 2007) designed and built a **Moon Jelly Strobilation** display for her Senior Project. She was able to successfully strobilate jellies which will ensure that we have a steady supply of adult jellies for our popular **kreisel**, or jelly display. So next time you stop by the Aquarium make sure to check out Lisa Scott's juvenile jellies in the new **Moon Jelly Strobilation Display.**

Note the number of technical terms in this short text (which I have highlighted here in boldface) with which *strobilation* collocates here and in the other citations below. And notice the semantic tension in a case like *jelly*, which, in its primary sense ('wobbly foodstuff made with gelatin'), is not a highly technical term, whereas in the sense in which it is used in this document, it clearly it is. This noun has two quite different meanings, with different sets of collocates. (ODE only has *jellyfish* for this sense.)

According to Wikipedia, the meaning of *strobilation* is as follows:

Strobilation or **transverse fission** is a form of asexual reproduction consisting of the spontaneous transverse segmentation of the body. It is observed in certain **cnidarians** and **helminths**. This mode of reproduction is characterized by high offspring output, which, in the case of the parasitic tapeworms, is of great significance.

The word is in *Merriam Webster's Online Dictionary*, with an entry as follows:

> Main Entry: **stro·bi·la·tion** ...
> Etymology: New Latin *strobila*
> **:** asexual reproduction (as in various coelenterates and tapeworms) by transverse division of the body into segments which develop into separate individuals, zooids, or proglottids.

The etymology could and should (but does not) go on to say that the New Latin term is based on Greek *strobilē* 'an act or state of twisting or whirling', which is related to *strobilos* 'pine cone' (ultimately a derivative of *strephein* 'to twist or whirl').

Collins English Dictionary, 5th edition on line (2000), has a clearer definition ('asexual reproduction by division into segments, as in tapeworms and jellyfishes') but does not offer an etymology. This is surprising, not least because the web page for this entry invites us to 'see also **strobile, strobilus, stroboscope'**. The intention is praiseworthy, but the connection between a *strobilus* (a pine cone), *stroboscopes* (which gave us *strobe lighting*), and *strobilation* must be baffling rather than enlightening to anyone who does not already know the etymology.

The phraseological preferences of terminological words are rarely of much interest. The collocation of *strobilation* and *jelly* helps to disambiguate *jelly* and adds to our understanding of what happens in the world, but does not contribute very much to our understanding of language in general or the meaning of the word *strobilation* in particular. Two other things are of interest in a case such as this: 1) the relation between the word and the world outside language; 2) the convoluted etymological history that takes us from pine-cones to asexual reproduction by way of the act of twisting.

3. The Phraseological Extreme

Let us now turn to the meaning of a much more ordinary, everyday word. What does the verb *blow* mean? Is it something that the wind does or something that an electrical apparatus or fuse does? Or is it something that mammals do? Or something to do with explosions? The answer, of course, is that this word is used to form phrases that activate all of these meanings and many more. The meaning is context-dependent. Corpus analysis can show us the contextual norms of phraseology that play a large part in determining the different meanings and activating their implicatures.

The essential point, for a lexical analyst, is to identify the various phraseological patterns or norms with which each such word is associated and to identify the meaning or implicature of each pattern, rather than to start by trying to dream up a list of meanings and then finding (or inventing) phraseological examples to illustrate the different meanings. The reason why the latter procedure is unreliable is that intuitions are an unreliable source of evidence. When we consult our intuitions in an attempt to discover the meanings of a polysemous word, we human beings all too often overlook some senses while giving overdue or distorted emphasis to others.

Only slightly less unreliable is the traditional procedure adopted by pre-corpus scholarly dictionaries such as OED and Merriam-Webster. This involves basing definitions on empirical evidence in the form of citations collected by readers. Here, the problem is one of overgeneralization on the basis of insufficient evidence. Citation readers tend to collect evidence for rare and unusual words and meanings – words with a strong terminological

tendency. It is not possible to measure the statistical significance of collocations on the basis of a collection of citations gathered by human readers. But measuring the statistical significance of collocations is essential for decoding phraseological meaning. A more reliable procedure is to discover, through painstaking analysis of large corpus samples, the patterns of valency and collocation in which the target word normally participates and then to make generalizations about the meaning of the patterns (or the word in context), rather than about the word in isolation.

Having said that, let us now return to phraseological meaning and our questions about the meaning of *blow*. Most words exhibiting a strong phraseological tendency are common and widespread and form part of the core of a language, with several different meaning potentials. The verb *blow* is the base for six phrasal verbs (*blow apart, blow away, blow down, blow off, blow over, blow up*), with several different meanings (*blowing up a balloon* activates a different meaning of *blow up* from that of *blowing up a building*). Moreover, this meaning is normally but not necessarily associated with a phrasal verb. In certain contexts, the same sense can be activated, as in 1, without the completive particle *up*, or with the completive phrase *to pieces*, as in 2.

1. The Germans had blown the bridges over the Po.
2. Gandhi walked towards a platform to begin his speech and was blown to pieces by a bomb.

Some but not all of the phrasal verbs associated with *blow* have causative/ inchoative alternations (someone may *blow up* a building, or the building may *blow up* of its own accord, for example due to a gas leak). Just three core terminological meanings can be identified for the base verb *blow*: what the wind does; what a person does by exhaling breath with some force (e.g. *blowing smoke over one's fellow diners, blowing bubbles,* or *blowing one's nose*); and what an explosion does (for example, *blow a hole in a wall*). Each of these core meanings, as also each phrasal verb, is associated both with a set of alternations (active/passive, causative/inchoative, conative, resultative) and with a set of collocational preferences (e.g. *wind, gale, sand, snow, ship; fire, nose, bubble; building, house, window, fuse*).

The verb *blow* is also associated with at least 17 idiomatic expressions, some of which overlap phraseologically with more literal meanings: e.g. an electrical appliance may *blow a fuse*, which is literal, but if a person *blows a fuse* (in the sense of losing their temper) we have a conventional idiom. Idioms with special meanings of their own include *blow* [a project] *off course*, *blow the cobwebs away* [= introduce fresh thinking], *blow a person away* [= shoot to death], *blow one's own trumpet* [= boast], *blow the whistle on someone or something* [= expose wrongdoing], *blow someone's cover* [= expose their disguise], *blow one's or someone else's brains out* [= kill], *blow a kiss to someone* (concrete action, nearly literal), *blow hot and cold* [= vacillate], *blow a proposition apart* or *to pieces* [= invalidate it], *blow a theory or a plan* [= invalidate it], *blow a fuse* [= lose one's temper], *blow one's stack* [= lose one's temper], *blow one's top* [= lose one's temper], *blow someone's mind* [= cause them to be amazed], *blow the gaff* [= expose a secret], and *blow a raspberry* [= make a rude, derisive noise]. All of these patterns and more for the verb *blow* are described in the on-line, freely available *Pattern Dictionary of English Verbs* (http://nlp.fi.muni.cz/projects/cpa). The analytic procedure of corpus pattern analysis that creates the entries in the *Pattern Dictionary* is described in Hanks (2004).

The task for the compiler of an electronic phraseological dictionary is to triangulate all the various phraseological constructions, idioms, and collocational preferences, and come up with

a reasonably clear, coherent, distinctive, and exhaustive set of clear statements relating meaning to use − statement, not of necessary conditions of meaning, but rather of typical implicatures. Some currently available dictionaries, e.g. Cobuild, can be seen to be struggling hard to represent this complex mass of phraseological data: others simply ignore it. Altogether, over 50 distinctive phraseological patterns for this verb are identified. For the patterns to be distinctive, valency alone is insufficient, and collocation alone is likewise insufficient. It is necessary to map onto valencies of the kind identified by Herbst et al. (2004) sets of collocations identified by semantic type, as proposed by Hanks and Pustejovsky (2005). Distinctive patterns of meaningful linguistic behaviour then emerge and can be recorded.

One other analytic ingredient is required for this enterprise, namely to distinguish normal usage from abnormal usage such as innovative metaphors. Idiomatic and phraseologically correct but abnormal uses of words are classified as exploitations of patterns of normal usage. A simple example is 3, from a text describing problems of birth control education in Nepal.

> 3.   When a visiting paramedic distributed free condoms, the children blew them up and played with them like balloons.

It is possible but not normal to blow up condoms. 3, therefore, is an exploitation of the normal pattern, *blow up a balloon.* Such uses are fascinating, but in analytic lexicography, thy must be set aside for separate analysis, because to take account of all such variations would interfere with our recognition of the underlying stereotypical pattern.

4. Words exhibiting tension between the phraseological tendency and the terminological tendency

Two examples may be given of words whose phraseological tendency clashes with the terminological tendency: the noun *second* and the verb *organize*.

In the case of *second*, a terminological definition has been stipulated by an international committee of scientists. This is not compatible conflicts with ordinary usage. The technical definition of second is:

> the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium 133 atom.

This definition was agreed by a committee of scientists that meets occasionally in Paris to stipulate terminology and definitions for the Système International (SI) of weights and measures. It is, no doubt, very important for use in scientific research contexts where very precise measurements are important. However, it has nothing to do with the use of this noun in everyday language, where the word is found in phraseology such as *wait a second, a split second, I was on my feet a second later, put your guitar down for a second, some of them take a second or two to stop whirring*, and so on. Here, the meaning is 'a very short period of time - much shorter than a minute', but seeking scientific precision of definition would obviously be irrelevant. The task of the general-language lexicographer is to find a way of reconciling the technical meaning and the everyday meaning, relating the latter to normal phraseology, while taking account of accepted terminological specifications, and making a concise meaningful presentation of the whole to the public.

5. Words in which the terminological tendency is complemented by the phraseological tendency

There is a problem with the verb *organize*, lexicographically speaking. It has a large variety of collocational preferences in the direct-object position and most of them entail very different kinds of action implied by the verb. It is hard to group them by any coherent semantic or pragmatic criterion. The lexical sets blur into one another in a puzzling way, so it is difficult to choose the right level of lexicographical generalization or to decide how many different meanings to report. Examples 4-12 illustrate the cline from one prototype to another.

4. To coincide with Gorbachev's arrival in Vilnius, Sajudis had organized a mass rally in the city's Cathedral Square.
5. Various radical pro-reform groups organized a huge demonstration in the centre of Moscow.
6. Attempts by the United States to organize a Middle East peace conference remained deadlocked.
7. Accommodation can be organized in hotels or with families.
8. Ideas and concepts held by people … are the historical product of the need to organize society so that human beings in society can produce and reproduce.
9. In about 1813-14 he organized in Bermondsey a society of workmen ….
10. Notation or code may be used in organizing books on shelves or files in a filing cabinet
11. All the evidence has led us to organize this book with reference to different kinds of places.
12. The brain is organized with a considerable amount of parallel wiring.

Prototypically, what is organized is an event, and some cognitive agent does the organizing. In BNC, according to the Sketch Engine, the dozen nouns in the direct object slot most associated with *organize* (*rally, demonstration, march, petition, strike, boycott, excursion, resistance, referendum, protest, coup, campaign*) are all indisputably events, as illustrated by examples 4, 5, and 6. The pattern seems to be quite simply:

[[Human]] organize [[Event]]

It is not a great problem to fit 7 into this pattern, since the noun *accommodation* is ambiguous: it can either denote the location in which a person stays temporarily, or (as here) the event of staying in it. However, it is less plausible to argue that 8-12 fit into this pattern. Neither society in general (8) nor particular societies (in the sense 'social group', as in 9) are events, and the same is true of books as physical objects (10), books as sources of information (11), and the brain (12). The brain is not an event. There is, in fact, an indefinite number of prototypical examples: *organizing an event* such as a demonstration or a peace conversation entails quite a different set of actions from *organizing the chapters in a book* that one is writing, and *organizing books on a shelf* is different again. Can a contrast be drawn between arranging an event such as a peace conference or a coup on the one hand and organizing an entity such as a book, a library, a society, or a brain on the other? The answer seems to be no. The first grouping is quite satisfactory, but the second is not. If a noun occurring in the direct-object slot in relation to *organize* can be identifies as having the semantic type [[Event]], it activates one and only one meaning of the verb, namely 'do whatever is necessary to ensure that [[Event]] takes place'. However, the same cannot be said of nouns having the semantic type [[Entity]]. They are heterogeneous, and activate heterogeneous meanings of the verb. Moreover, with some entities, an organizing agent is present, while with others it is not. Organizing books in a library, chapters in a book, and social groups are activities undertaken by humans as cognitive agents, but there is no cognitive agent (other than, perhaps, God) that determines the way that the brain is organized or the way in which the feathers are arranged in a peacock's tail. This (usually passive) use of *organize* has more to do with reporting perceptions of patterns in the world than with the activities of a cognitive agent. Thus, the

evidence for some normal uses of this word defies using the semantic type of the direct object as a contribution to analysing the meaning of the verb.

Yet another different conventional sense (or pair of senses) of *organize* is represented by 13 and 14.

    13. Ticha was one of the most active leaders in Aguilares. FECCAS had formed a branch there and she helped **organize** women coffee harvesters.
    14. Agricultural workers were still denied the right to **organize**.

Women coffee harvesters are, of course, entities not events. It is undoubtedly true that women, being human beings, are entities - but 13 does not imply that that Ticha (whoever she was) imposed a neat, patterned arrangement on the women referred to while they did their coffee harvesting. It implies quite specifically that she organized them as a labour union in order to secure improvements in their pay and conditions as workers. This political sense of sense of *organize* is the only one that has a regular nul-object alternation. Conventionally, when the verb *organize* is used intransitively (inchoatively), as in 14. it is the trade-union sense that is activated. It would be perfectly possible to say to a chaotic person with a mass of papers around him or her, '*You really should organize!*' but in fact, we do not say this. If someone did say this, they might well be interpreted as recommending trade-union activities.

What is going on here? There seems to be one very general sense of *organize*: something is put (or appears to have been put) into some kind of coherent order. This general sense is largely independent of phraseology. On the other hand, some very specific norms are activated is particular collocations. Thus, terminologically specific senses of *organize* appear to be dependent on phraseology, while the general sense, which we might expect to be able to classify as terminological, is independent of phraseology. This contrasts with the case of second and seems to be incompatible with Sinclair's prediction. The two tendencies seem to be independent variables.

Thus, verbs such as *organize* present a lexicographical dilemma, namely whether to report one sense, three senses, or half a dozen. One can state just one very general sense, with a number of constraints (e.g. this verb is not normally used intransitively), but there are a few rather specific uses that are more or less closely related to the general sense. It would appear to be optional whether to report these domain-specific subsenses There is no single right answer to this kind of dilemma. The best that can be aimed at in a dictionary is arbitrary imposition of consistency of treatment of such words across all the affected entries in any one dictionary. In a lexical semantic account, it is necessary to explain the prototype and its variations.

Another example in which the phraseological tendency complements the terminological tendency is the verb *admit*. There are just two general meanings for *admit*, which may be summarized as:

A) to say reluctantly
B) to allow to enter

Each of these terminological meanings is associated with a distinctive set of phraseological patterns with more specific implicatures. For example, in a sample of uses of *admit* in the 'say reluctantly' sense, taken from BNC, over 50% have as the direct object or clausal object a word or phrase denoting some bad action performed by the subject of the sentence. However,

it is important to note that this is not a necessary condition of this sense of *admit*. Contrast 15 and 16.

15. Richmond **admitted** driving a motor vehicle with excess alcohol in his blood.
16. Mary Bryce **admits** a certain sympathy for Norman Lamont.

Driving a motor vehicle with excess alcohol in one's blood (15), is according to all the social conventions of the civilized world, a bad thing to do. However, 16 is more complex. It implies that, as far as Mary Bryce is concerned, feeling sympathy for Norman Lamont (a former Conservative politician and chancellor of the exchequer) is or ought to be a bad thing to do. This is not necessarily the view of Mary Bryce herself. 16 tells us explicitly that Mary Bryce felt sympathy for Normal Lamont, but it also tells us something about the speaker's attitude. By choosing the verb *admits*, the speaker implies that someone - possibly the speaker, possibly the subject of the sentence, Mary Bryce - believes that feeling sympathy for Normal Lamont is a bad thing. Speech-act verbs (among others) are riddled with implicatures about the attitidue of the speaker, the attitude attributed to the subject or object of a sentence, and the attitudes attirbitued by the speaker to society as a whole.

Now let us look at two examples of verb use in which the terminological meaning is closely related to some event in the world, but largely independent of the phraseology.

17. A sluice on the lock is opened to admit water.
18. The skylights in the galleries **admit** light through angled screens.

17 and 18 are perfectly well-formed and meaningful uses of the verb *admit*. They activate terminological sense B, 'to allow to enter' at its most literal. However, admitting stuff such as light and water are comparatively rare uses of this verb. 17 and 18 are terminologically normal, but phraseologically exceptional, precisely because they are comparatively rare. This particular sense of *admit* ('allow to enter') is almost always activated in one of a number of specific phraseologies that imply not merely 'allow to enter' but also activate some domain-specific (terminological) procedure. If a human or a human group is admitted to something, then that something is normally some kind of organized social grouping, with specific rule-governed procedures for admission. These are illustrated in examples 19-21, each of which implies a different formal admission procedure.

19. Each old person **admitted** to residential care should sign a contract relating to the rights of residents.
20. They [the Baltic nations] were **admitted** into the United Nations and other international organizations.
21. Joanna had dislocated her hip and was **admitted** to hospital.

19 in particular may be singled out for further discussion, as it shows well how collocation of two polysemous words activates just one relevant and highly specific sense of each of them, entailing in addition a highly specific set of implicatures. The noun *care* is polysemous, but if a person is admitted to care, the implication is that the person in question is classified as one who is unable to attend to his or her own needs and is in need of protection. The kind of care provided is that offered by social services, not just any kind of caring, and certainly not the group of senses of *care* that imply trouble or worry.

6. Conclusions

•	The meaning potential of a word consists of a puzzling mixture of terminology and phraseology: contextual dependence and contextual independence. Terminological specificity

and phraseological boundedness are independent variables.

- Particular phraseological contexts often activate highly specific terminological meanings, with domain-specific implicatures (as in the case of *admit someone to care; organize the workers*).

- In other cases (e.g. *strobilation, second*) the terminological tendency consists of a relation between the term itself and some aspect of the world rather than of language and is not much affected by phraseology.

- Much modern technical terminology is of Latin and Greek origin, and this is relevant to understanding its meaning. On-line dictionaries can and should afford the space to pay careful attention to explaining etymons clearly. This is particularly important in the case of words like the *strob-* group, where an Ancient Greek word has spawned a wide variety of different meanings in modern languages and the semantic relations between them (e.g. between asexual reproduction in tapeworms and strobe lights in a disco) are obscure to the uninitiated.

- At the other extreme, the discovery procedures for phraseological norms requires detailed corpus analysis of much authentic data. The account of *blow, v.,* discussed above, is based on painstaking analysis of 1500 concordance lines from BNC.

- Many words exhibit both a phraseological tendency and a terminological tendency. The two tendencies may be in harmony, as in the case of *admit*, where specific phraseology elaborates the general terminological tendency, or there may be some kind of tension between them, as in the case of *second* and *organize*.

References

Hanks, Patrick. (1994). 'Linguistic norms and pragmatic explanations, or why lexicographers need prototype theory and vice versa'. In F. Kiefer, G. Kiss, and J. Pajzs (eds.). *Papers in Computational Lexicography: Complex '94*. Research Institute for Linguistics, Hungarian Academy of Sciences.

Hanks, Patrick. (2004). 'Corpus Pattern Analysis'. In: G. Williams and S. Vessier (eds.), *Euralex Proceedings*. Vol. 1. Lorient: Université de Bretagne-Sud.

Hanks, Patrick, and James Pustejovsky. (2005). 'A pattern dictionary for natural language processing'. In *Revue Française de Linguistique Appliquée* 10 (2).

Herbst, Thomas et al. (2004). *A Valency Dictionary of English: a corpus-based analysis of the complementation patterns of English verbs, nouns and adjectives*. Berlin: de Gruyter

Kilgarriff, Adam. (1997). 'I don't believe in word senses'. In *Computers and the Humanities* 31 (2).

Sinclair, John. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Sinclair, John. (1998). 'The lexical item'. In E. Weigand (ed.). *Contrastive Lexical Semantics*. John Benjamins.

Wray, Alison. (2002). *Formulaic Language and the Lexicon*. Cambridge University Press.