# Elliptical Arguments: a Problem in relating Meaning to Use

Hanks, Patrick
Institute of Formal and Applied Linguistics, Charles University, Prague
patrick.w.hanks@gmail.com

Electronic dictionaries of the future will be much in demand—for computational, pedagogical, and other applications—if they can be used as resources for mapping word meaning systematically onto word use. Research in computational linguistics and artificial intelligence over the past twenty years, despite many declarations of success, has shown that existing dictionaries, designed for human users, coupled with existing linguistic theory of a top-down, speculative nature, have failed to be suitable for this goal. Nor are results using hierarchical ontologies such as WordNet any better. Such resources are very plausible for human users, but they fail to meet the challenges of mapping meaning systematically onto words in use in ordinary text. In the English-speaking world, vast sums of funding have been poured into computational linguistic research on the 'Word Sense Disambiguation problem', using dictionaries for foreign learners such as LDOCE as a resource, with what some of the protagonists (e.g. Ide and Wilks 2006) now acknowledge are disappointing results. There are several reasons for this, chief among them being the fact that the research question has been formulated on false assumptions. The chief of these is what Fillmore (1976) charaterized as a "checklist theory of meaning". The implication that words have a finite and discrete list of senses that can be disambiguated by some procedure or other is a dangerously crude generalization, encouraged by superficial inspection of traditional dictionaries, but not consistent with the evidence of word use in a large corpus. This paper attempts to show why, and what should be done instead.

Insufficient attention has been paid to the highly variable nature of word usage in ordinary unsupervised text. In this paper, I propose to show that such variations are rule-governed exploitations of norms, which electronic dictionaries of the future will be obliged to make explicit. This implies moving away from 'Lego-set' theories of language of a Fregean variety, in which words are put together like children's toy bricks in order to make meaningful propositions. Instead, effective future electronic dictionaries will have to give an account of the normal phraseology associated with each word and how meanings are associated with phraseological patterns, including patterns in which certain arguments are not explicitly realized at all. Thus, a lexical entry word in an e-Dictionary will be no more than an index item, an entry point to an inventory of normal phraseological patterns with which meanings and/or translations and/or other implicatures are associated. This requires a distinction between normal patterns of word use and abnormal uses, which are exploitations of norms.

Although word usage in everyday texts is highly variable, variation is not random. Corpus analysis, using tools such as Sketch Engine (Kilgarriff et al., 2004), shows that, underlying the many variations, usage is highly patterned. Moreover, the variations themselves constitute sets of secondary patterns. A simple example will illustrate the point. Consider the verb *fire*. Over a dozen different patterns of normal use of this verb can be distinguished by corpus analysis. Some of these patterns activate very similar

meanings; others activate quite different meanings. In the most basic pattern, exemplified in (1), the meaning is 'discharge a projectile from a firearm'. This contrasts with other meanings of the same verb, e.g. (2), 'to stimulate or excite', (3), 'to bake in a kiln', and (4) 'to dismiss from employment'.

(1)  I was in a place once when a man fired a gun at me and I did not like it at all.[1]

(2)  Active citizenship has already fired the imagination of many people.

(3)  Fashioning and firing a pot does not affect the clay composition.

(4)  This time General Avril fired four lieutenant-colonels.

In these examples, the semantic types of the arguments activate different senses of the verb. In (1), the direct object is a firearm, in (2) it is a psychological entity, in (3) it is a pot, and in (4) it is a person.  Each of these direct objects correlates with the semantic types of other arguments; for example, in (1) and (4), there is a correlation with the subject, which is normally a word deonting person, but (1) also correlates with and adverbial of direction ('at me'), which (4) does not. The direct object in (2) typically governs a dependent possessive ('of many people', '*our*') and typically correlates with a subject denoting an abstract entity.

It would be very convenient if natural language <u>always</u> behaved in the way suggested by these carefully selected examples of normal patterns of use. However it does not. A pattern dictionary must not only discover and describe phraseological patterns in a way that has not yet been done adequately by any dictionary; it must also be accompanied by a theory (or a rule book) describing the rules that govern variations in each pattern. One such variation is ellipsis (i.e. omission). In ordinary language use,  there are some circumstances in which an argument can be omitted, while in other cases  they cannot. These omissions rarely bother human readers and hearers, because the speaker or writer correctly judges the omitted item to be 'obvious'. The ellided argument is taken to  be common kn owledge and therefore not worth stating. Perhaps for this reason, ellipsis has not been adequately described in lingustic theory.  However, it can be a very serious problem for computers and language learners alike.  It is hard enough to write NLP programs that process data in text; processing data that is not presetn in the text but that is 'understood' poses an extra challenge.  Electronic dictionaries of the future must therefore account for the precise circumstances under which ellipsis is possible.

With (1), both the direct object and the adverbial of direction are optional. One can say:

(1a)  *I was in a place once when a man fired at me and I did not like it at all*

or:

(1b)  *I was in a place once when a man fired a gun and I did not like it at all.*

In an appropriate context, one can even say

(1c) *He fired*

and mean he fired a gun. What's more, if the verb is intransitive with no adverbial, the meaning <u>must</u> be that he fired a gun and not that he dismissed someone from employment, nor that he baked a pot in a kiln, nor that his ideas inspired enthusiasm in

1   In this paper, the convention is followed of printing 'real' examples (taken from corpora—BNC unless otherwise stated— and other texts) in roman, while invented examples (used mainly for contrastive purposes) are in italics.

others. Thus, 1c is quite unambiguous, even though only one argument is explicitly realized. A further complication, as FrameNet shows, is that there is yet another semantic type competing for the direct object slot in 1, namely the projectile. If it is present, it drives out the semantic type [[Fiream]] from this slot, either completely, as in (1d) or into an adverbial slot (as in 1e).

(1d) *A man fired several rounds at me.*

(1e) *A man fired several rounds at me from a revolver.*

By contrast with (1), the patterns illustrated in (2), (3), and (4) do not allow optional omission of arguments. You cannot say '*he fired*' and mean that he dismissed somebody. You cannot say, '*\*Active citizenship fired many people*' and mean that it inspired them with enthusiasm. In fact, the latter is ungrammatical as well as invented.

If a particular argument type is added to pattern (4), a different sense is activated, or rather an ambiguity arises, requiring additional context to distinguish the meaning.

(4a) *General Avril fired four lieutenant-colonels with enthusiasm.*

(4a) can mean either that General Avril took a sadistic delight in dismissing the lieutenant-colonels or that his ideas inspired them with positive feelings. The wording is identical but the clause structure of each of the two interpretations is quite different. In the first interpretation, the adverbial is an optional adjunct; in the second interpretation, it is a necessary part of the pattern, with an ambiguity resulting from metonymy in the subject: the term **General Avril** (the whole) is being used as metonym for a part or attribute of General Avril (his ideas or policies).

This all sounds very complicated. Does it imply that pattern-based lexicography is impossible? The answer is a resounding no! Pattern-based lexicography is perfectly possible, but firstly we have to reconstruct the theoretical basis of lexicographical practice, and secondly we have to learn to ask the right questions when analysing data. We also have to resist both the temptation to invent examples that fit our preconceived notions about what each word means and the computational linguist's yearning for magic bullets and algoithms that will solve all problems at a single stroke. And we have to be equally sceptical about checklists of word meanings (attached to words in isolation, out of context), as found in almost all current dictionaries. Words in isolation have meaning potential, not meaning, as argued in Hanks (1994, 2000) and elsewhere.

Practical considerations—from the point of view of both lexicographers and dictionary users—suggest that, if a comprehensive account is to be given of the relationship between words and meanings, a two-part structure for lexical analysis of any word is necessary. First, the normal patterns of use must be identified and a robust account given of the meaning of each pattern. Depending on the syntagmatic characteristics of each verb and its position on the cline between idiomaticity and open choice (Sinclair, 1991), this will account for between 70% and 90% of all uses of most verbs. To account for the remaining uses, a set of exploitation rules must be invoked. One such rule is ellipsis. The dictionary entry for each word must specify precisely the circumstances under which ellipsis—and every other kind of exploitation—is possible.

The presentation concludes with sample entries, constructed on the principles explained here and by Hanks and Pustejovsky (2005), from the *Pattern Dictionary*

*of English Verbs (PDEV; http://nlp.fi.muni.cz/projects/cpa/*). These entries are supported by evidence from BNC and, where appropriate, linked to FrameNet.

**References**

Fillmore, Charles J. (1975). 'An alternative to checklist theories of meaning'. In *Papers from the First Annual Meeting of the Berkeley Linguistics Society.* Reprinted in Hanks (ed., 2008c), *Lexicology: Critical Concepts in Linguistics,* volume 4..

Hanks, P. (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. Reprinted in Hanks (ed., 2008c)*,* volume 5.

Hanks, P. (2000). 'Do word meanings exist?'. In *Computers and the Humanities* 34. Reprinted in T. Fontenelle (ed., 2008), *Practical Lexicography: a Reade*r. Oxford University Press.

Hanks, P. (2008a). 'Why the "Word Sense Disambiguation Problem" can't be solved, and what should be done instead". In: B. Lewandowska-Tomaszczyk (ed.), *Corpus Linguistics, Computer Tools, and Applications: State of the Art*. Peter Lang.

Hanks, P. (2008b). 'The Lexicographical legacy of John Sinclair'. In *International Journal of Lexicography*, 21:3.

Hanks, P. (ed., 2008c). *Lexicology: Critical Concepts in Linguistics,* 6 volumes. Routledge.

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

Ide, N., and Wilks, Y. (2006). 'Making sense about sense'. In: E. Agirre and P. Edmonds (eds.) *Word Sense Disambiguation: Algorithms and Applications.* Springer.

Kilgarriff, A., Rychlý, P., Smrž, P., and Tugwell, D. (2004). 'The Sketch Engine' in *Proceedings of Euralex.* Lorient, France. Reprinted in Hanks (ed. 2008c), volume 6.

Pustejovsky, J. and Hanks, P. (2001). 'Very large lexical databases'. ACL Tutorial, Toulouse.

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

**Acknowledgement**