

- Herold, A. (2005a): *Suchanfragen: Dokumentation zur Korpusabfrage und zur Arbeit mit der Idiomdatenbank*. Internal ms., Berlin-Brandenburg Academy of Sciences.
- Herold, A. (2005b): Reducing the Size of Sample Corpora for Research on Idioms in the German Language. Poster presented at the Conference on Corpus Linguistics, University of Birmingham, UK, July 2005.
- Kramer, U., Neumann, G., Stathi, K. and Fellbaum, C. (2005): Kollokationen im Wörterbuch. Das Wolfgang Paul-Preis Projekt an der Berlin-Brandenburgischen Akademie der Wissenschaften. *Zeitschrift für Germanistik*.
- Lezius, W. (2000): Morphy - German Morphology, Part-of-Speech Tagging and Applications. In Heid, U., Evert, S., Lehmann, E. and Christian Rohrer (eds), *Proceedings of the 9th EURALEX International Congress*. Stuttgart, Germany, 619-623.
- Moon, R. (1998a): Frequencies and Forms of Phrasal Lexemes in English. In Cowie, A. (ed.), *Phraseology: Theory, Analysis, Applications*. Oxford: Oxford University Press, 79-100.
- Moon, R. (1998b): Fixed Expressions and Idioms in English: A Corpus-based Approach, (Oxford Studies in Lexicography and Lexicology). Oxford, Oxford University Press.
- Neumann, G., Fellbaum, C., Geyken, A., Herold, A., Hümmer, C., Körner, F., Kramer, U., Krell, K., Sokirko, A., Stantcheva, D., and Stathi, K. (2004): A Corpus-Based Lexical Resource of German Idioms. In Saint Dizier, P. & Zock, M. (eds) *Proceedings of the Workshop on Electronic Lexicons*, COLING, Geneva, 48-52.
- Nunberg, J., Sag, I. & Wasow, T. (1994): Idioms. *Language*, 70, 491-538.
- Schemann, H. (1993): *Deutsche Idiomatik. Die deutschen Redewendungen im Kontext*. Stuttgart and Dresden.
- Sokirko, A. (2003): DDC - A Search Engine For Linguistically Annotated Corpora. *Proceedings of Dialogue 2003*, Protvino, Russia, June 2003.
- Sokirko A. (2004): Morphological components on www.aot.ru. *Proceedings of Dialogue 2004*, Russia, Verchnevolzhskiy.

A Pattern Dictionary for Natural Language Processing

Patrick Hanks and James Pustejovsky
Brandeis University

Abstract: *This paper briefly surveys three of the main resources for word sense disambiguation that are currently in use – WordNet, FrameNet, and Levin classes – and proposes an alternative approach, focusing on verbs and their valencies. This new approach does not attempt to account for all possible uses of a verb, but rather all its normal uses ('norms'). By corpus pattern analysis (CPA), the normal patterns of use of verbs are established. A meaning ('primary implicature') is associated with each pattern. The patterns are then available as benchmarks against which the probable meaning of any sentence can be measured. The status of abnormal or unusual uses ('exploitations') is also briefly discussed. Also, three kinds of alternation are recognized: syntactic diathesis alternations, semantic-type alternations, and lexical alternations.*

Résumé : *Cet article passe en revue de façon succincte trois des ressources principales utilisées actuellement pour la désambiguïsation lexicale (WordNet, FrameNet et les classes de Levin), et propose une approche alternative, en prenant comme point de départ les verbes et leurs valences. Cette nouvelle approche ne tente pas de rendre compte de tous les usages possibles d'un mot, mais plutôt de tous ses usages normaux (les 'normes'). Les patrons normaux d'utilisation des verbes sont dégagés par une méthode que nous appelons Corpus Pattern Analysis (CPA, analyse des patrons basée sur les corpus). A chaque patron se trouve associé un sens (une 'implication principale'). Les patrons sont ensuite utilisés comme des étalons par rapport auxquels on peut mesurer le sens probable de n'importe quelle phrase. Nous abordons aussi le statut des usages anormaux ou inhabituels (les 'exploitations'). Nous reconnaissons aussi trois types d'alternance : les alternances syntaxiques liées à la diathèse, les alternances de type sémantique et les alternances lexicales.*

1. Overview: Lexical Resources

For a wide variety of NLP applications, a lexicon with information about how words are used and what they mean is a necessary component. Pustejovsky (1995) shows how even limited amounts of default context associated with a lexical item can offer major improvements in the compositional operations associated with natural language systems. In this paper, we illustrate an alternative, more radical approach.

Lexical resources currently available include WordNet, FrameNet, and Levin Classes, each of which has its strengths and its weaknesses. We comment briefly on the salient characteristics of each and show why a new, empirically well-founded resource, with criteria for distinguishing one sense of a word from another, is both necessary and possible. Specifically, such a resource will assign stereotypical semantic values and roles to the valencies of each verb for each of its senses. These stereotypical semantic values and roles play a large part in distinguishing the different senses of a verb in context.

In the Appendix, we present three entries from the “Corpus Pattern Analysis” (CPA) project currently being compiled at Brandeis University. The aim of CPA is to link word use to word meaning in a machine-tractable way. Words in isolation, we have found, do not have specific meanings; rather they have a multifaceted *potential* to contribute to the meaning of an utterance. Different facets of this potential are realized in different contexts. Corpus evidence shows that contextual patterns of word use are very regular, although abnormal contexts also occur, sometimes accidentally, but more often for rhetorical effect. For this reason, attempts to account for all possible meanings of a word are misguided. Projects with this aim tend to produce impractical results, because normal usage becomes buried in a welter of remote possibilities. Our goal is more limited and more practical: it is to account for all normal meanings of each word.

Local context is usually sufficient to assign a specific sense to a word and to distinguish one sense from another. Discovering the normal contexts in which words are used reduces lexical entropy dramatically. We classify abnormal contexts (such as those created by poets) as exploitations of norms. CPA discovers the normal patterns, sets aside exploitations and other oddities, and attaches a meaning (a ‘primary implicature’) to each normal pattern.

The focus is on verbs. For CPA, the entry point to a sentence is its verb. Large samples of actual uses of each verb are taken from a corpus (the British National Corpus, BNC), as described in Hanks (2004). The valencies are analysed and semantic values (**types** and **roles**) are assigned to each valency. A semantic type is a class to which a term can be assigned, e.g. *Peter* or *the old man* belong to the semantic type [[Person]]. In the context of treating patients, Peter or the old man is acting as a doctor or other health professional; whereas in the context of being treated by a doctor, Peter or the old man fulfils the role of patient. These are context-specific roles. **Semantic roles** are linked to **semantic types** in CPA by an equals sign, thus: [[Person=Doctor]], [[Person=Patient]]. (There is, of course, a lot more to semantic typing than this, but in the limited space available here this will give a general idea of what we do.)

The result will be a dictionary of normal sentence patterns in English, to which hitherto unseen sentences in free text can be matched for assignment of a meaning or for any of various other NLP purposes. CPA links word use to word meaning in a hard-nosed, empirically testable way. It provides a checklist, not as a set of necessary conditions that must be met, but rather as a set of contextual benchmarks against which the likely meaning of any given utterance can be measured.

When applied to previously unseen text, CPA matching is a powerful and subtle tool, but of course it depends on and interacts with other analytic processes, including word-class tagging, parsing, pronoun anaphora resolution, and semantic typing. If any of these are wrong in a given sentence, then the results of CPA matching are unpredictable. A positive aspect of this is that CPA can contribute to the improvement of such resources, e.g. to parsers by highlighting recurrent parsing errors and to anaphora processors by indicating the likely semantic class of a pronoun’s antecedent.

2. Available Disambiguation Resources

Three main resources are commonly cited in the literature.

2.1. WordNet

The great merit of WordNet (Fellbaum, 1998) is that it is a full inventory of English words (along with a number of terms such as *craniate* and *chordate* which are found neither in ordinary English nor in ordinary scientific discourse in the relevant subject, but which seem rather to be taxonomically motivated terms invented to fill a node in a semantic hierarchy). WordNet assigns words to “synsets” (synonym sets), which are equated with “senses”. Specifically, according to WordNet’s on-line glossary, a *sense* is “a meaning of a word in WordNet. Each sense of a word is in a different *synset*.”

Members of the NLP community seem to have accepted with little or no discussion WordNet’s equation of synsets with senses. Closer inspection, however, shows that many of WordNet’s senses are indistinguishable from one another by any criterion – syntactic, syntagmatic, or semantic – other than the fact that they happen to have been placed in different synsets. For example, in WordNet 2.1 the verb *write* is said to have 10 senses:

1. **write**, compose, pen, indite – (produce a literary work; *She composed a poem; He wrote four novels*)
2. **write** – (communicate or express by writing; *Please write to me every week*)
3. publish, **write** – (have (one’s written work) issued for publication; *How many books did Georges Simenon write?; She published 25 books during her long career*)
4. **write**, drop a line – (communicate (with) in writing; *Write her soon, please!*)
5. **write** – (communicate by letter; *He wrote that he would be coming soon*)
6. compose, **write** – (write music; *Beethoven composed nine symphonies*)
7. **write** – (mark or trace on a surface; *The artist wrote Chinese characters on a big piece of white paper*)
8. **write** – (record data on a computer; *boot-up instructions are written on the hard disk*)
9. spell, **write** – (write or name the letters that comprise the conventionally accepted form of (a word or part of a word); *He spelled the word wrong in this letter*)

10. **write** (create code, write a computer program); *She writes code faster than anybody else.*

These are hardly different senses, but rather different facets of a single sense or (as in the case of 1 and 3) repetitions of exactly the same sense, associated with different synonyms. The arguments of Fillmore (1975) against “check-list theories of meaning”, Pustejovsky (1995) against a “sense-enumerative lexicon” (one that enumerates different facets of the same sense as separate senses), and Wierzbicka’s advice to lexicographers to “seek the invariant” (Wierzbicka, 1993, 51-57) are relevant here.

WordNet’s synsets are built into a gigantic hierarchical ontology. Do the nodes in this hierarchy represent semantic classes and do those classes fulfill particular slots in verb argument structure? Examination of the superordinates (hyperonyms) of each synset suggests that the answer has to be No. In many places, WordNet’s hierarchies and distinctions do not correspond to anything empirically observable. They are figments of the compiler’s imagination, sometimes plausible, sometimes less so. Thus, the superordinates of the ten synsets containing *write* in WordNet are given as:

1. create verbally
2. communicate, intercommunicate
3. create verbally
4. correspond
5. create verbally
6. make, create (which is itself a superordinate of ‘create verbally’)
7. trace, draw, line, describe, delineate
8. record, tape
9. [No superordinate].
10. create code, write a computer program

Even if the hierarchy of semantic types were to be pared down and reorganized – as they have been in EuroWordNet (Vossen, 1998) – the nodes in the hierarchy, with their current populations of words, often fail to generate the words needed to express a syntagmatic pattern. For this reason, CPA often specifies a lexical set (see “LEXSET” in the sample entries below) extensionally, by simply enumerating typical members. In such cases it is often an open question whether *any* semantic feature unifies the relevant lexical items into a node in a semantic hierarchy. In other cases it is obvious that an intensional semantic property such as [[Human]] or [[Artefact]] is the only sensible way in which a large lexical set can be populated.

2.2. FrameNet

Fillmore’s work in case grammar and frame semantics is justifiably famous and does not need to be recapitulated here. It is full of insight and, among other things, serves as a reminder of the holistic nature of verb argument structure, with alternations in the syntactic slots in which a particular semantic argument may be realized. FrameNet (Atkins et al., 2003, Fillmore et al., 2003, Ruppenhofer et al., 2005) aims

roles implied by the semantics of each word are both stated and exemplified explicitly (regardless of whether they necessarily occur in all sentences in which the word is used). For example, if someone *risks* their life or their wealth, a desirable goal is implied, whether or not it is explicitly mentioned in any given utterance.

FrameNet uses corpus data extensively, but it proceeds frame by frame, not word by word. It relies on the intuitions of its researchers to populate each frame with words. This runs the risk of accidental omissions, and it means that (in principle) no word can be regarded as completely analysed until all frames are complete. At the time of writing, there has been no indication of when that will be, nor of the total number of frames that there will be. Currently, some frames overlap to the point of being indistinguishable (see comments on *fire* below). Others are only partly populated. Unfortunately, some frames announce a lexical entry as complete, when in fact only minor or rare senses have been covered. For example, the verb *spoil* is currently a member of two frames in FrameNet: **Rotting** and **Desiring**. **Rotting** is the ‘rotting meat’ sense, which may be cognitively salient but is actually quite rare. The **Desiring** frame is exemplified in the phrase ‘spoiling for a fight’. Together, these two senses account for less than 3% of all uses of this verb in BNC. The main uses (‘spoil an event’ and ‘spoil a child’) are not yet covered.

If CPA succeeds in its objective of analysing all the normal uses of each verb, it will complement FrameNet neatly in this respect. FrameNet offers a very full and detailed semantic analysis of each frame; CPA offers a contrastive analysis of the senses of each word. When a CPA entry for a given verb is finished, it has, by definition, completed analysis of all normal uses of that verb.

2.3. Levin Classes

The first part of Levin (1993) discusses diathesis alternations of verbs. The notion of alternations is a useful one for CPA. Some of the alternations discussed (e.g. causative/inchoative; unexpressed object) are pervasive in English, though others are rare.

CPA adds the concept of a semantic alternation to that of a syntactic diathesis alternation. For example, for the medical sense of *treat*, the lexical set [[Person=Doctor]] alternates with [[Medicament]], while in the direct object slot [[Person=Patient]] alternates with [[Injury]] and [[Ailment]]. In cases such as this, two or more different semantic types in a given valency pick out the same sense of the verb. There is also lexical alternation, as in *grasping/clutching at straws*, where the words may alternate without any change in the basic meaning.

In the second half of the book Levin attempts a classification of some English verbs based on her own intuitions about their meaning, supported by the intuitions of other academics who have written about them. Levin argues that the behaviour of a verb is to a large extent determined by its meaning. It could equally well be argued that the meaning of a verb is to a large extent determined by its behaviour. This seems to be a chicken-or-egg question and therefore unanswerable – or rather, the answer may be no more than a matter of taste and theoretical preference. There is, however, a practical reason for taking analysis of a word’s typical behaviour as a

starting point for analysis, rather than its meaning. Word behaviour is observable and verifiable by inspection of recurrent uses in large corpora, search engines, etc., whereas a word's meaning is imponderable, a matter of introspection, conjecture, and unsubstantiated assertion. In monolingual lexicography, there are well-established guidelines (varying slightly from dictionary to dictionary) for supporting each definition with examples of actual usage and for cross-checking the actual wording of definitions with other team members, in order to guard against highly idiosyncratic interpretations. Levin classes do not seem to have been compiled with the benefit of any such safeguards or cross-checks. Many of Levin's assertions about the behaviour (and sometimes also the meaning) of particular verbs in her verb classes are idiosyncratic or simply wrong. Our findings accord with those of Baker and Ruppenhofer (2002), that when compared with actual usage, Levin's comments about diathesis alternations for verb classes apply to some but not all members of the classes. This is a pervasive problem in the second half of the book. Detailed examples are given below.

As a matter of practicality, Levin deliberately excludes verbs that take sentential complements from her research. For this reason, *tempt* is listed only as an "Amuse verb" (31.1); there is no mention of its more normal use with a sentential complement, as in *We were tempted to laugh*. Levin discusses approximately 3,000 English verbs. She does not by any means cover all of the major verbs (no entry for *specialize*, *specify*, *spell*, *spend*, *spoil*, etc., although some much rarer verbs such as *spellbind* are included), nor – more significantly for purposes of word-sense disambiguation – does she cover all of the major senses of the verbs that she does include. It therefore comes as something of a surprise to find that, some twelve years after their publication, Levin classes are widely cited in the NLP community as if they had some sort of established empirical validity. This may be taken as evidence of the hunger of the research community for some resource, any resource, however limited, that links meaning and use.

3. Supplementary Clues

The combination of the semantic values of the valencies (subject, object, and what may be dubbed 'argumental adverbial') assigns a distinctive basic sense to verbs in use. For example, *fire a gun* (= cause to discharge a bullet) contrasts with *fire a person* (=dismiss from employment). More subtly, CPA also distinguishes *fire a gun* from *fire a bullet from a gun*, which is necessary if NLP is going to recognize that bullets are not guns. But sometimes more information is needed. For example, *shoot a person* could conceivably be ambiguous, depending on whether the subject of the sentence is an armed attacker or a film director. Thus, the semantic role of the subject of *shoot* in turn assigns a semantic role to the direct object. If the subject is an armed attacker and the direct object is a person, then the direct object is a victim. If the subject of the sentence is a film director, the direct object is an actor. However, the information as to whether the person is an armed attacker or a film director may not be available. Therefore, CPA specifies not only the semantic type of the typical arguments of a verb (its valencies), but also additional relevant and recurrent clues if any. For example, *shoot a person dead* and *shoot and injure a*

person are common expressions that are quite unambiguous, so the resultative adjective *dead* and the coordinated verb *injure* are noted in CPA as supplementary clues in the relevant pattern.

Likewise, the pattern "[[Person]] *gallop* [Adv[Direction]]" is not ambiguous at a basic level, insofar as it implies swift movement and resonates with the more literal sense "[[Horse]] *gallop*". However, it is ambiguous insofar as it may be **metonymic** – the person in question may be a rider on horseback – or **metaphorical** – the person in question may be on foot. In some but not all cases, the [Adv[Direction]] provides a disambiguating clue. (If the person gallops into a hotel and up the stairs, he or she is probably not on horseback.)

Thus, the sense of a verb in context is built up in explicit detail on the basis of such contextual clues as represent normal usage. CPA records a central group of such clues for each verb. CPA also records the comparative frequency of each pattern in the training data; this could provide a basis for default interpretations in cases of uncertain matches.

4. Conclusion

In this short paper, we have critically examined three of the major lexical resources available in the field. There are other important efforts that we have not discussed, however. For example, much work has been done using electronic versions of print dictionaries that were originally compiled for human users (see, for example, Stevenson & Wilks, 2003). For NLP purposes, the main problem with such dictionaries is that they do not show explicitly how the meanings they describe can be mapped onto actual usage. Other frequently cited resources include VerbNet (Palmer *et al.*, 2004), PropBank (Palmer *et al.*, 2004), and NomBank (Meyers *et al.*, 2005). Because of their inherent dependence on WordNet and Levin classes, much of our criticism above is applicable to aspects of these resources as well. Still, even with such criticism, it is important to recognize how valuable the development of these resources has been for the community. Our goal in this paper has been to demonstrate how the CPA methodology can substantially improve the coverage, accuracy, and utility of lexically encoded contexts.

CPA is slowly and painstakingly building up an inventory of normal syntagmatic behaviour that may be useful for word sense disambiguation, message understanding, natural text generation, and other applications. The approach is illustrated in the three CPA entries in the Appendix. Having established the template and procedures for CPA, our next step must be to scale up. A lexicographer has laboriously compiled entries for just over 100 verbs. Altogether the English language contains approximately 8000 verbs, of which approximately 6000 have more than one sense according to the *Concise Oxford Dictionary*. Compiling a pattern dictionary for 6000 or more verbs will involve substantial effort. We are encouraged in this effort by the results of automatic lexical set clustering and induction as reported in Pustejovsky *et al.* (2004).

Patrick Hanks and James Pustejovsky
 Department of Computer Science
 Brandeis University
 Waltham, MA 02454, USA
 hanks@bbaw.de; jamesp@cs.brandeis.edu

References

- Atkins, S., Rundell, M. and Sato, H. (2003): The Contribution of FrameNet to Practical Lexicography. *International Journal of Lexicography*, 16:1, 333-357.
- Baker, C. and Ruppenhofer, J. (2002): FrameNet's Frames vs. Levin's Verb Classes. In J. Larson and M. Paster (eds.), *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*, 27-38.
- Fellbaum, C. (ed.) (1998): *WordNet: An Electronic Lexical Database*. Cambridge (MA), MIT Press.
- Fillmore, C.J. (1975): An Alternative to Checklist Theories of Meaning. In Cogen, C. et al. (eds), *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, Berkeley (CA), BLS, 123-131.
- Fillmore, C.J., Johnson, C. and Petruck, M.R.L. (2003): Background to FrameNet. *International Journal of Lexicography*, 16-1, 235-250.
- 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, 89-114.
- Hanks, P. (1996): Contextual Dependency and Lexical Sets. *International Journal of Corpus Linguistics* 1(1), 75-98.
- Hanks, P. (2004): Corpus Pattern Analysis. In Williams, G. & Vessier, S. (eds), *Euralex Proceedings*. Vol. I, Lorient, France, Université de Bretagne-Sud, 87-98.
- Levin, B. (1993): *English Verb Classes and Alternations: a Preliminary Investigation*. University of Chicago Press.
- Macleod, C., Grishman, R. and Meyers, A. (1998). *COMLEX Syntax Reference Manual*. Proteus Project, NYU. Comlex is distributed through the Linguistic Data Consortium (LDC98L21).
- Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, B. and Grishman, R. (2004). The NomBank Project: An Interim Report. *Proceedings of the HLT-EACL Workshop on Frontiers in Corpus Annotation*, Boston (MA).
- Palmer, M., Gildea, D., Kingsbury, P. (2005). The Proposition Bank: A Corpus Annotated with Semantic Roles, *Computational Linguistics Journal*, 31-1.
- Pustejovsky, J. (1995): *The Generative Lexicon*. Cambridge (MA), MIT Press.
- Pustejovsky, J., Rumshisky, A. and Hanks, P. (2004): Automated Induction of Sense in Context. Geneva, *COLING 2004 Proceedings*.
- Pustejovsky, J., Meyers, A., Palmer, M. and Poesio, M. (2005). Merging PropBank, NomBank, TimeBank, Penn Discourse Treebank, and Coreference. *ACL 2005 Proceedings of Workshop on Frontiers in Corpus Annotation II*, Ann Arbor.
- Ruppenhofer, J., Ellsworth, M., Petruck, M.R.L. and Johnson, C.R. (2005). *FrameNet: Theory and Practice*. On-line publication at <http://framenet.icsi.berkeley.edu/>
- Vossen, P. (1998): Introduction to EuroWordNet. *Computers and the Humanities*, 32, 73-89.

- Wierzbicka, A. (1993). What's the Use of Theoretical Lexicography? *Dictionaries: Journal of the Dictionary Society of North America*, 14, 44-78.
- Stevenson, M. & Wilks, Y. (2003): Word Sense Disambiguation. In R. Mitkov (ed.) *The Oxford Handbook of Computational Linguistics*. Oxford University Press, 249-265.

Websites

- FrameNet. <http://framenet.icsi.berkeley.edu/>
 WordNet 2.1. <http://wordnet.princeton.edu/>

APPENDIX

THREE CPA VERB ENTRIES, WITH COMMENTARY

1. GRASP

The verb *grasp* has 3 senses and 8 patterns in CPA.
 There is a conative alternation.
 There are 2 idioms.

COMMENTARY

Grasp typically denotes the act of seizing something rather the state of holding something. The main semantic split is between grasping a physical object and grasping an idea. Grasping an idea could be classified as a metaphorical exploitation of the physical-object sense, but it is a very frequent conventional expression, accounting for nearly two thirds of all uses in BNC.

A split is also made in CPA between grasping a physical object and grasping a person, but this split is very fine. Patterns 2 and 4 could easily be lumped together. A person is, after all, a physical object. On the other hand, lumping them would make it impossible to attach different implicatures to these two patterns. For this reason they have (provisionally) been kept separate.

A conative alternation (patterns 3 and 5) is found for both physical and mental objects. This alternation is instantiated by the prepositions *at* and *for*.

The sense of *grasping an opportunity* (pattern 6) is sometimes lumped together with *grasping a concept* (pattern 4), but semantically they are quite distinct.

Continuous aspect (*to be grasping something*) is rare, and normally occurs only with physical, not mental objects. With a physical object, the sense is affected by the aspect: *to grasp something* or *to have grasped something* implies an action, but *to be grasping something* implies a state.

The idiom *grasp the nettle* is a Britishism. The idiom *grasp at straws* is a variant of *clutch at straws*. Its sense is conative.

grasp: CPA and WORDNET

The verb *grasp* is found in two synsets in WordNet, which correspond to the two main uses of the verb (represented in CPA as patterns 1, 2, and 4):

1. grasp, hold on.
2. get the picture, comprehend, savvy, dig, grasp, compass, apprehend.

WordNet does not cover CPA patterns 3 and 5 (conative alternations), 6 (*grasp an opportunity*), 7 (the British idiom *grasp the nettle*) or 8 (*grasping at straws*).

grasp: CPA and FRAMENET

FrameNet has a **Grasp** frame, which it defines as follows:

A **Cognizer** possesses knowledge about the workings, significance, or meaning of an idea or object, which we call **Phenomenon**, and is able to make predictions about the behavior or occurrence of the Phenomenon. The Phenomenon may be incorporated into the wider knowledge structure via categorization, which can be indicated by the mention of a **Category**. The Cognizer may possess knowledge only in part and this may be expressed in a **Completeness** expression. Note that the knowledge may have been acquired either from instruction or from the Cognizer's own experimentation, observation, or mental operations.

Words in this frame are frequently used metonymically to denote the transition into the state described above.

Grasp is also in the **Manipulation** frame, which is defined thus:

The words in this frame describe the manipulation of an **Entity** by an **Agent**.

There is no mention in FrameNet of the conative alternation, nor of the sense 'seize an opportunity' (CPA pattern 6).

grasp: CPA and LEVIN CLASSES

Levin classifies *grasp* as a "Hold verb" (15.2), along with *clasp*, *clutch*, *grip*, *handle*, *hold*, and *wield*.

Levin asterisks the conative alternation for class 15.2, indicating that she thinks these verbs do not participate in it. Against this, there is good evidence in BNC that *clutch*, *grasp*, and *clasp* – though not *grip*, *handle*, *hold*, or *wield* – are sometimes used conatively, for example "her hands were grasping at his coat"; "the goalkeeper was left clutching at thin air"; "people clutched at the coffin as it was carried to the graveyard"; "My hands close around his neck; his own hands involuntarily rise to clasp at my fingers".

CPA patterns show that *grasp* is more frequently a verb of seizing rather than of holding. Levin places *seize* in two classes, neither of which seem appropriate for *grasp*: 1) as a "verb of possessional deprivation" like *steal* (10.5), and 2) as an "obtain verb" with benefactive alternation (13.5).

Levin makes no mention of the 'understand' senses of *grasp*, although this is in fact its most common use. There is no Levin class of verbs involving comprehension or understanding, presumably because these verbs sometimes take sentential complements.

GRASP: THE PATTERNS AND THEIR PRIMARY IMPLICATURES

I. SEIZE HOLD OF SOMETHING

1. [[Person]] **grasp** [[PhysObj]] (14%)

IMPLICATURE: [[Person=Animate]] seizes [[PhysObj]] and holds it firmly.

LEXICAL ALTERNATION: [[Person]] <-> {hand, finger}

OTHER CLUES: {in [POSDET] hand}, {by [DET] arm}

EX.: *He grasped the handle of the door in one hand, and that of the spoon in the other.*

He reached out wildly, trying to grasp the creature, but it had moved away.

2. [[Person 1]] **grasp** {[[Person 2]] (by [[BodyPart | Clothing]])} | {[POSDET] [[BodyPart | Clothing]]} (13%)

IMPLICATURE: [[Person 1=Animate]] seizes [[BodyPart]] or [[Clothing]] of [[Person 2=Animate]]

LEXICAL ALTERNATION: [[Person 1]] <-> {hand, finger}

EX.: *The defender moves forward and grasps the attacker's leg.*

Benjamin stretched across and grasped the man's hand.

Laura grasped Maggie by the arm.

3. [[Person]] **grasp** [NO OBJ] {at | for} [[PhysObj]] (2%)

IMPLICATURE: [[Person=Animate]] attempts to seize [[PhysObj]].

COMMENT: conative alternation of 1 and 2.

EX.: *Theda had gone paler than usual, and she grasped at the bedpost for support.*

The child was still crying as Alan sat down with him, but he grasped greedily for the milk.

II. UNDERSTAND SOMETHING

4. [[Person]] **grasp** {[[Abstract]] | [N-clause]} (59%)

IMPLICATURE: [[Person=Cognitive]] understands {[[Abstract=Concept]] | [N-clause]}

CLUES: easy to grasp, simple to grasp, hard to grasp, difficult to grasp.

EX.: *I know it did, but sometimes I can't grasp the reality.*

In the end we will grasp the truth.

I was too intelligent not to be already grasping the rules of the game we played.

After fifteen minutes or so, Julia thought that she had grasped most of the story.

He could never grasp the essentials, the requirements, the obligations of living in a western society.

Teachers should grasp the fact that the DES can lay down details of a policy but that the Department of Employment funds it.

He had not grasped that Ruby worked that day with a mere photograph.

She grasped what was happening.

5. [[Person]] **grasp** [NO OBJ] {at [[Abstract]]} (<1%)

IMPLICATURE: [[Person=Cognitive]] attempts to understand or call to mind

[[Abstract=Concept]].

COMMENT: conative alternation of 4.

EX.: In this Jarman sits, Prospero-like, sniffing flowers as if grasping at a memory of happier times.

III. USE AND OPPORTUNITY

6. [[Person]] **grasp** [[LEXSET Opportunity]] (5%)

IMPLICATURE: [[Person]] takes advantage of [[Opportunity]].

LEXSET [[Opportunity<Abstract]]: opportunity, chance, offer, moment ...

EX.: *Lawrence hoped his players would grasp the chance of cup glory.*

The Prime Minister failed to grasp that opportunity.

Kylie, singing like she had never before, grasped the moment.

IDIOMS

7. [[Person]] **grasp {the nettle}** (3%)

IMPLICATURE: [[Person=Cognitive]] takes bold and decisive action to deal with a problem.

COMMENT: Idiom.

EX.: *Ian Corner, David Chell and their staff are bravely grasping the nettle of recession.*

The Labour Party has failed to grasp the nettle in Monklands.

And that's what the G M B need to do, to grasp the nettle to move forward.

8. [[Person=Cognitive]] **grasp {at {straw}}** (1.5%)

IMPLICATURE: [[Person]], facing a desperate situation, attempts to implement {straw = a solution that has little chance of success}.

COMMENT: Idiom. Variant of 'clutch at straws' (see CLUTCH)

EX.: *Nadirpur's eyes widened. He was grasping at straws.*

2. SEEK

The verb *seek* has 8 patterns in CPA:

- 1 phrasal verb (1 pattern);
- 2 main verb senses (5 patterns);
- 2 idiomatic phrases (2 patterns).

COMMENTARY

The majority of uses of this verb all have the same general sense: to try to find or get something. The underlying semantics involve both a physical object and an event – the event of finding the sought object. In fact, the pattern of seeking an event (e.g. *seek a compromise*) is more common than seeking a person or physical object.

There is a great deal of semantic coercion: the precise purpose and/or means of seeking are determined by a combination of the verb and its direct object: e.g. a lawyer *seeking the death penalty* asks a court to impose it on a criminal; a band

seeking a drummer advertises; an entrepreneur *seeking funds* makes a proposal to venture capitalists, and so on.

When *seek* is transitive, the direct object slot can have a wide range of semantic values: people seek *actions, ideas, money, other people, physical objects*, etc. CPA divides these into three classes: 1) [[Event | Abstract]]; 2) [[Money]]; 3) [[Person | PhysObj]]. [[Event | Abstract]] is selected as the main semantic value for the direct object of *seek* as this is more frequent than other semantic values.

seek: CPA and WORDNET

The verb *seek* has 5 “senses” in WordNet.

1. **seek** – (try to get or reach; '*seek a position*'; '*seek an education*'; '*seek happiness*')
2. **search, seek, look for** – (try to locate or discover, or try to establish the existence of; '*The police are searching for clues*'; '*They are searching for the missing man in the entire county*')
3. **try, seek, attempt, essay, assay** – (make an effort or attempt; '*He tried to shake off his fears*'; '*The infant had essayed a few wobbly steps*'; '*The police attempted to stop the thief*'; '*He sought to improve himself*'; '*She always seeks to do good in the world*')
4. **seek** – (go to or towards; '*a liquid seeks its own level*')
5. **seek** – (inquire for; '*seek directions from a local*')

CPA Pattern 1 maps in part onto the WordNet phrasal verb *seek out*.

CPA Pattern 2 maps roughly onto WordNet Synsets 1 and 5.

CPA Pattern 3 has no equivalent in WordNet.

CPA Pattern 4 maps roughly onto WordNet Synset 2.

CPA Pattern 5 also maps roughly onto WordNet Synset 2.

CPA Pattern 6 maps roughly onto WordNet Synset 3.

WordNet Synset 4 offers a distinction that is not justified, according to CPA.

seek: CPA and FRAMENET

FrameNet has *seek* in only one frame, the **Seeking** frame, which is defined as follows:

A **Cognizer_agent** attempts to find some **Sought_entity** by examining some **Ground**.
The success or failure of this activity (the **Outcome**) may be indicated.

FrameNet does not have *seek* in the **Attempt** frame. CPA Pattern 6 suggests that this may be an oversight.

seek: CPA and LEVIN CLASSES

Levin classifies *seek* as a “ferret verb” under “Verbs of Searching” (section 35). The “ferret verbs” are: *ferret, nose, seek, and tease*. It is hard to see any good reason for classifying these verbs together. The normal use and meaning of each of them is quite different from the meaning assigned to them by Levin. For three of them (but not for *seek*) a completive adverbial is obligatory, but Levin does not mention this.

a) *Ferret* is a verb of finding rather than a verb of searching, and as such it normally takes the completive-intensive particle *out*.

b) *Nose* is only a verb of searching when it is intransitive with the completive adverbial *around* or a prepositional phrase governed by *through* or a similar preposition. As a transitive verb with *out*, it is a verb of finding (not of searching), and is quite rare.

c) *Tease* is not a verb of searching. The nearest equivalent is *tease something out*, which means 'establish the facts about something', not 'search for something'. In this case, too, *out* is obligatory. *Tease* minus *out* means something completely different.

d) *Seek* is a verb of searching, but (contrary to Levin's assertion, p. 199), *seek something* sometimes alternates with *seek for something*.

SEEK: THE PATTERNS AND THEIR PRIMARY IMPLICATURES

PHRASAL VERB: SEEK OUT

1. [[Person | Animal]] **seek** [[Entity]] {out} (4%)

IMPLICATURE: [[Person | Animal]] looks for and finds [[Entity]].

IMPLICATURE 2: Perseverance may be needed to find [[Entity]].

COMMENT: Completive-intensive variant of pattern 4.

EX.: *Bragg and Morton went through the unpretentious entrance of the New Club ... and sought out the Secretary.*

Tour operator Airtours has set up a hit-squad of inspectors to seek out and axe shoddy continental hotels.

We will seek out people to put on the database.

Fish normally seek out the warmer shallows.

MAIN VERB

I. TRY TO GET OR FIND SOMETHING

2. [[Person]] **seek** [[Event | Abstract]] (63%)

IMPLICATURE: [[Person=Cognitive]] attempts to activate [[Event | Abstract]]

EX.: *It's best to try to seek a compromise rather a perfect solution.*

At Bodycare '92 you will be able to seek advice from the experts.

With his house plunged into darkness, Jessie's owner sought my help.

The question could be thought to arise of whether they are seeking revenge.

Young adults, themselves seeking a separate sexual and personal identity, ...

When a person is seeking employment,

3. [[Person]] **seek** [[Money]] (4%)

IMPLICATURE: [[Person]] tries to get [[Money]], typically as compensation for an injustice or funding for a project.

LEXSET [[Money<Entity]]: compensation, damages, grant, payment, ...

EX.: *Farmers are being urged to seek grants to help restore hundreds of miles of stone walls in North Yorkshire.*

In many cases the problem could be solved by seeking an interim payment.

4. [[Person 1 | Animal]] **seek** [[PhysObj | Stuff | Person 2]] (10%)

IMPLICATURE: [[Person 1 | Animal]] wants to find [[PhysObj | Stuff | Person 2]]

COMMENT: This is a catch-all ragbag category for anything not covered by 2 or 3.

EX.: *Christian Rock band seeks drummer in Surrey area.*

African magazine seeks local writers.

Any reader seeking an individual document at a library...

.. a French agent sought by New Zealand in the bombing (with one death) of the Greenpeace ship Rainbow Warrior .

5. [[Person]] **seek** [NO OBJ] {for [[TopType]]} (<1%)

IMPLICATURE: [[Person]] tries to find [[TopType]]

COMMENT: This is a variant of patterns 2-4. Never passive.

EX.: *But none of this was in Sara's mind as she patted Jenny's arm, seeking for some words of comfort.*

Much of the early science of plant ecology sought for correlations between vegetation and physical, not biotic, factors in the environment.

II. TRY TO DO SOMETHING

6. [[Person]] **seek** [to/INF V] (18%)

IMPLICATURE: [[Person]] attempts to do [V]

EX.: *She sought to hide her nervousness with levity.*

The Purchaser seeks to have the best of both worlds.

...an isolated settlement popular with tourists seeking to experience the African bush in the wild.

IDIOM

7. [[TopType]] **be** {not far to seek} (<1%)

IMPLICATURE: [[TopType]] is easy to find or call to mind.

EX.: *The reasons are not far to seek.*

PAST PARTICIPLE AS PREDICATIVE ADJECTIVE: 'SOUGHT AFTER' (<1%)

8. [[TopType]] **be** {sought after}

IMPLICATURE: [[TopType]] is very popular and much in demand.

EX.: *York ham commands a high price and is much sought after.*

3. FIRE

The verb **fire** has 8 senses and 15 patterns in CPA.

COMMENTARY

fire: CPA and WORDNET

The verb *fire* is a happy example where, in broad terms, WordNet and CPA can be seen to be describing the same language. There are nevertheless some significant

differences of detail.

The verb *fire* has 9 “senses” in WordNet.

1. open fire, fire – (start firing a weapon)
2. fire, discharge – (cause to go off; ‘fire a gun’; ‘fire a bullet’)
3. fire – (bake in a kiln so as to harden; ‘fire pottery’)
4. fire, give notice, can, dismiss, give the axe, send away, sack, force out, give the sack, terminate – (terminate the employment of; ‘The boss fired his secretary today’; ‘The company terminated 25% of its workers’)
5. fire, discharge, go off – (go off or discharge; ‘The gun fired’)
6. fire – (drive out or away by or as if by fire; ‘The soldiers were fired’; ‘Surrender fires the cold skepticism’)
7. arouse, elicit, enkindle, kindle, evoke, fire, raise, provoke – (call forth (emotions, feelings, and responses); ‘arouse pity’; ‘raise a smile’; ‘evoke sympathy’)
8. burn, fire, burn down – (destroy by fire; ‘They burned the house and his diaries’)
9. fuel, fire – (provide with fuel; ‘Oil fires the furnace’)

CPA Patterns 1 (causative) and 3 (unexpressed object) map onto WordNet senses 1 and 2. However, CPA does not accept the WordNet distinction between firing and beginning to fire as separate meanings of the verb *fire*; this distinction is made in English by other means, e.g. *open fire vs. continue firing*. It is therefore hard to justify the distinction between WordNet senses 1 and 2.

CPA Pattern 2 maps partly onto WordNet sense 2. However, WordNet collapses the distinction between firing a gun and firing a bullet from a gun, which is of some importance for NLP, since a gun is not a projectile.

CPA Pattern 4 maps onto WordNet sense 5.

CPA Pattern 6 maps onto WordNet sense 4.

CPA Pattern 9 maps onto WordNet sense 9.

CPA Pattern 10 maps onto WordNet sense 3.

CPA Pattern 11 maps roughly onto WordNet sense 8.

CPA Pattern 12 maps onto WordNet sense 7.

CPA Patterns 5, 7, 8, 13, 14, and 15 are not covered by WordNet.

WordNet sense 6 is incoherent and seems unnecessary.

fire: CPA and FRAMENET

The verb *fire* is given in three FrameNet frames.

1. **Shoot_Projectiles**: An **Agent** causes a **Firearm** to discharge which induces the rapid motion of the **Projectile**.
2. **Use Firearm**: An **Agent** causes a **Firearm** to discharge, usually directing the projectile from the barrel of the **Firearm** (the **Source**), along a **Path**, and to a **Goal**.
3. **Firing**: An **Employer** ends an employment relationship with an **Employee**. There is often a **Reason** given for the action.

The distinction between Frames 1 and 2 is so subtle that it is hard to see any justification for it. CPA patterns 1-4 are all no more than common syntactic alternations of the sense given in these two frames, which would be better accounted for in a single frame.

CPA pattern 6 maps onto the **Firing** frame.

The remaining CPA patterns (5, 7, 8, 9, 10, 11, 12, 13, 14, 15) are not covered by FrameNet. FrameNet makes no mention of *firing questions at people*, *firing machines up*, *firing a boiler with fuel*, *firing pots in a kiln*, or *firing people with enthusiasm*. With one exception, these are minor senses. The exception is that 11% of uses of the verb *fire* have the sense ‘inspire with enthusiasm’ (as in *It was not only her matching handbag and high-heeled shoes which fired my enthusiasm*). FrameNet has a **Subjective_Influence** frame which mentions *inspire*, but not *fire* (*with*).

fire: CPA and LEVIN CLASSES

Levin categorizes *fire* as a ‘Throw Verb’ (17.1), i.e. ‘a verb of instantaneously causing ballistic motion by imparting a force’. Her only other mention of *fire* is at 2.8, the ‘with/against’ alternation, which she illustrates with the example *Brian hit the stick against the fence/Brian hit the fence with the stick*. The idiomaticity of *Brian hit the stick against the fence* is highly questionable. It may or may not be possible, but it is not natural. Apparently we are supposed to conclude that *Brian fired the gun/bullet against the target* and *Brian fired the target with the gun/bullet* are well-formed sentences of English and that they have the same meaning. There is no empirical evidence to support either conclusion, and they are at variance with our intuitions as native speakers of English.

Levin makes no mention anywhere of the alternation *fire a gun/fire a bullet from a gun*. And of course she makes no mention of prepositional phrases headed by *at*, which complement *fire* far more often than *with* or *against*. Describing normal phraseology is not part of Levin’s agenda. In short, only one sense of *fire* is covered by Levin classes, and even that disregards normal usage. Levin makes no mention of the other seven senses identified by CPA.

FIRE: THE PATTERNS AND THEIR PRIMARY IMPLICATURES

I. DISCHARGE A PROJECTILE FROM A GUN AT A TARGET

1. [[Person]] **fire** [[LEXSET Firearm]] (at [[PhysObj]]) (11%)

IMPLICATURE: [[Person]] causes [[Firearm]] to discharge a projectile toward [[PhysObj=Target]]

COMMENT: Often passive.

LEXSET [[Firearm<Artifact]]: gun, pistol, revolver, rifle, cannon, mortar, blunderbuss, weapon,

EX.: *Holmes and I fired our revolvers together.*

The terrorists were obviously about to fire their weapon.

He was interrupted as the gun was fired three times.

Mr Sinclair said the gun had only been fired once.

A second soldier burst through the door ... and fired his automatic rifle at the kitchen table.

2. [[Person]] **fire** [[LEXSET Projectile]] (off) ({from [[LEXSET Firearm]])} ({at [[PhysObj]]} | [ADV [Direction]]) (26%)

IMPLICATURE: [[Person]] causes [[Firearm]] to discharge [[Projectile]] toward [[PhysObj=Target]]

COMMENT: Often passive.

LEXSET [[Projectile<Artifact]]: bullet, round, shell, shot, volley, flare, rocket, blast, burst, salvo, broadside, barrage, torpedo, grenade, missile, Exocet, blank, (Verey */N), ...

LEXSET [[Firearm<Artifact]]: See Pattern 1.

EX.: *But at Stillington a shot was fired from a 12-bore shotgun.*

He fired off a volley of shots from his semi-automatic rifle.

Each time a single shot was fired.

Loyalist terrorists fired a missile at the top security Crumlin Road jail in Belfast last night.

One man ... fired two shots from a handgun into the officer's chest.

3. [[Person]] **fire** [NO OBJ] ({at [[PhysObj]]} | {on [[HumanGroup]]} | [ADV [Direction]]) (20%)

IMPLICATURE: [[Person]] causes a gun or other firearm to discharge a projectile (in a given direction)

COMMENT: This is an 'unexpressed object' alternation of 1.

EX.: *He ordered his men to fire.*

He more or less admits that he fired first.

He was walking with his mum in Chicago when two men in a car began firing at three others.

The high court said Traore had ordered his forces to fire on demonstrators.

On Sept. 3 troops were also seen to fire on a crowd in Sebokeng.

Slash picked up the guy's rifle and began firing into the air.

4. [[LEXSET Firearm]] **fire** [NO OBJ] ({at [[PhysObj]]} | {on [[HumanGroup]]} | [Adv[Direction]]) (5%)

IMPLICATURE: [[Firearm]] discharges a projectile toward [[PhysObj=Target]] or [[HumanGroup=Target]]

COMMENT: Inchoative alternation of 1.

EX.: *The creature's weapon fired but missed her.*

A Barrett Light Fifty rifle, which fires a huge armour-piercing half inch diameter shell...

II. DIRECT QUESTIONS OR ACCUSATIONS AT SOMEONE

5. [[Person 1]] **fire** [[SpeechAct | Document]] ({at | to} [[Person 2]]) (3%)

IMPLICATURE: [[Person 1=Cognitive]] utter [[SpeechAct | Document]] to [[Person 2]]

IMPLICATURE 2: [[SpeechAct | Document]] is a question or an accusation.

EX.: *Volleys of accusations were fired from both sides.*

"What's the matter with you?" she fired at him furiously.

Each time a memo was fired off and the problem was solved.

III. DISMISS FROM EMPLOYMENT

6. [[Person 1]] **fire** [[Person 2]] (for [[Action=Bad]]) (11%)

IMPLICATURE: [[Person 1=Employer]] dismiss [[Person 2=Employee]] from employment

ADDITIONAL CLUES: get fired; be fired; [[Action]] get someone fired; you're fired

EX.: *Miguel Rafaelo could fire her tomorrow.*

He was fired for serious misconduct.

You're fired.

IV. START

7. [[Person]] **fire** [[Machine]] (up) (1%)

IMPLICATURE: [[Person]] causes [[Machine]] to start.

EX.: *She put her crash hat on and fired up the engine.*

Jack ... fired up his bike.

8. [[Machine]] **fire** [NO OBJ] (up) (2%)

IMPLICATURE: [[Machine]] starts.

COMMENT: Inchoative alternation of 7.

EX.: *The Jeep fired up and he slipped it into gear.*

Formerly there were glow-plugs which had to be warmed before the engine would fire.

9. [[LEXSET Fuel]] **fire** [[Machine]] (2%)

IMPLICATURE: [[Fuel]] causes [[Machine=Engine]] to operate.

LEXSET [[Fuel<Stuff]]: oil, coke, coal, wood, gas, ...

EX.: *The central heating was fired by oil.*

Furniture makers already use an estimated 35,000 tonnes of waste wood to fire boilers for space and water heating.

V. BAKE IN A KILN

10. [[Person]] **fire** [[LEXSET Pot | Brick]] (in {kiln}) (2%)

IMPLICATURE: [[Person]] bakes [[Pot | Brick]] in a kiln at high temperature in order to harden it and make it ready for use.

LEXSET [[Pot<Artifact]]: pot, pottery, porcelain, ...

LEXSET [[Brick<Artifact]]: brick, ...

EX.: *When fired in a special kiln, the metals fused onto the pot, giving the lustrous effect.*

When pottery is fired, the energy stored in the crystals is released as light.

VI. SET FIRE TO

11. [[Person]] **fire** [[Stuff | PhysObj]] (<1%)

IMPLICATURE: [[Person]] fire [Stuff | PhysObj]=Combustible]

COMMENT: Archaic, rare.

EX.: *Three constables went along with torches, firing the furze.*

Knots of horsemen ... were moving along the banks of the Tay, firing the jetties

that were not already broken.

VII. INSPIRE

12. [[TopType]] **fire** {[[Person]]'s [[LEXSET Enthusiasm]]} (7%)

IMPLICATURE: [[Person]] be filled with enthusiasm because of [[TopType]]

LEXSET [[Enthusiasm<Abstract]]: enthusiasm, imagination, interest, artistic and historical sense, heart, motivation

COMMENT: Often passive.

EX.: *My interest in robots was fired.*

It was not only her matching handbag and high-heeled shoes which fired my enthusiasm.

Her Majesty 'was extremely surprised at first, but then her artistic and historical sense was fired and she agreed'.

13. [[LEXSET Influence]] **fire** [[Person]] (up) (4%)

IMPLICATURE: [[LEXSET Influence]] causes [[Person]] to feel strongly.

LEXSET [[Influence<TopType]]: influence, anger, hatred, ambition, admiration; (& COERCIONS:) politics of conscience, principles of democracy, poetry, story, habit of contemplating, ritual movements, ...

COMMENT: Often passive.

EX.: *Anger, and the chance to shout it to the wind in the elemental language of their common childhood, fired her.*

By then, fired by an ambition to go all the way, he turned professional.

From her student days Pavlova was fired by stories of her legendary predecessor Marie Taglioni.

Modigliani ... had lost the "habit of contemplating landscape" that had fired him as a boy.

VIII. SPORTS SPECIFIC PATTERNS (the tip of an iceberg)

14. [[Person]] **fire** [[LEXSET Ball]] ([ADV [Direction]]) (4%)

IMPLICATURE: [[Person=Player]] kicks, hits, or passes [[Ball]] in a specific direction

DOMAIN: Sports journalism (Ball games)

LEXSET [[Ball<PhysObj]]: ball, winner, rebound, cross, goal, ...

EX.: *Stuart Gauld ... fired the ball low to the goalkeeper's right.*

Andy Nicol, bloodied from the fray, fires the ball from a ruck.

After a scrappy opening, Haslemere took the lead when James Simpson fired in a rebound.

15. [[Person]] **fire** [[HumanGroup]] {ahead | {into lead}} (2%)

IMPLICATURE: [[Person=Player]] causes [[HumanGroup=Team]] to take the lead by scoring a goal.

DOMAIN: Sports journalism (ball games)

EX.: *After Jürgen Klinsmann had fired the Germans ahead Riedle scored twice in five minutes.*

The Color of Things:

Towards the automatic acquisition of information for a descriptive dictionary

Gregory Grefenstette

LIC2M / CEA LIST, Fontenay-aux-Roses, France

Résumé : *Les dictionnaires contiennent souvent des descriptions des aspects visuels des objets. Mais l'information qui serait utile pour certaines applications dans le traitement de l'image y manque parfois. Nous présentons ici une méthode pour compléter ces descriptions par une analyse de texte du web, ainsi qu'une première analyse des résultats. Nous montrons que des informations concernant les couleurs typiques des objets peuvent être découvertes par une analyse purement textuelle et statistique.*

Abstract: *Physical objects are often described in dictionaries by visual features. But the information needed by computer applications for image analysis is not always found in dictionaries, nor in a complete form in any other publicly available information source. This article describes some first steps in finding more complete visual information about objects that could be used to enhance computer usable dictionaries and other knowledge repositories. We show that some information about the common colors of objects can be extracted automatically from text found on the Web.*

1. Introduction

It is tempting, with the great mass of text available via the Internet, to see whether it is possible to extend current dictionaries with automatically extracted information. We recently estimated that the amount of English text accessible through a popular search engine grew from 76 to 145 billion words from 2001 to 2004 (Grefenstette, 2004). With such volumes of text, there now exists the possibility of knowing what company a word keeps in printed language, on a large-scale. From exploiting this data to know what a word means (Firth, 1957) has been a dream of recent corpus linguists.

Traditional, descriptive dictionaries define the meaning of a word by other words presumed to be understood by the human reader. When a lexicographer creates a dictionary entry that describes a "cherry" as "any of numerous trees and shrubs producing a small fleshy round fruit with a single hard stone", the lexicographer supposes that the user knows what a "tree" and "shrub" are, as well as "fruit" and "stone", along with the other adjectives and verbs used there. In addition, there is the