



# **A Way with Words:**

**Recent Advances in Lexical Theory and Analysis  
A Festschrift for Patrick Hanks**

edited by Gilles-Maurice de Schryver



This landmark publication is a Festschrift for the famous corpus linguist and lexicographer **Patrick Hanks**, presented on the occasion of his 70th birthday. It is, however, far more than a dutiful collection of tributes. The 20 essays have a coherent theme – the relationship between word use and word meaning in the context of dictionary making and analysis of corpus evidence. These thought-provoking essays deal with major issues in every aspect of current thinking about the lexicon, some from a theoretical point of view, others from a more practical point of view.

This book is essential reading for everyone interested in meaning, the lexicon, dictionaries, and corpus analysis.

Dedicated to someone whose contribution to the field of corpus-based lexical analysis is second to none. — *Sue Atkins*

For the last thirty years, Patrick Hanks has been forging an account of how words and phrases work that reconciles the challenge of the philosophy with the detail. — *Adam Kilgariff & Pavel Rychlý*

I shall recall just one thing, a thing by which Patrick amazed me, as I am sure he has amazed many others since. What he showed me is how in English, whatever familiar verb you choose, if you set out a large and representative enough collection of contexts into which that verb will fit, then no other verb, however apparently similar in sense, will fit into all these contexts. Cumulatively, the complements for the verb add up to a unique pattern which rules out all candidates but one. — *David Wiggins*

Every lexicographer has their interests. I have had the opportunity to work with a number of different lexicographers. Many lexicographers study butterflies, but Patrick likes “central and typical” moths. — *Ken Church*

As a master of elegance, Johnson’s natural successor is Patrick Hanks. And if Hanks’s friends and colleagues can see more clearly, it is because we are standing on the shoulders of a giant. — *Michael Rundell*



[www.menhapublishers.com](http://www.menhapublishers.com)

ISBN 978-9970-10-101-6



9 789970 101016

Knihovna Mat.-fyz. fakulty  
Informatické oddělení  
Malostranské náměstí 25  
118 00 Praha 1

I 244/2010

Published in Uganda by  
Menha Publishers (U) Ltd.  
P.O. Box 35409 Kampala  
<http://www.menhapublishers.com/>

Branch Office  
Residentie Wellington  
F. Rooseveltlaan 381  
B-9000 Ghent (Belgium)

© Menha Publishers 2010

© John Sinclair 2007 (Chapter 2)

All photographs taken by  
Gilles-Maurice de Schryver

First published 2010

ISBN 978-9970-101-01-6

All rights reserved. No reproduction, copy or transmission of this publication may be made without the written permission from Menha Publishers. Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages.



6.	The Paradox of Analysis and the Paradox of Synonymy	
	<b>David Wiggins</b> .....	119
	<i>From 1994 until 2000, Wykeham Professor of Logic, University of Oxford</i>	
Part II: Computing Lexical Relations .....		133
7.	More is More	
	<b>Kenneth W. Church</b> .....	135
	<i>Johns Hopkins University</i>	
8.	Estimating the Number of Concepts	
	<b>Gregory Grefenstette</b> .....	143
	<i>Exalead</i>	
9.	Identifying Adjectives that Predict Noun Classes	
	<b>David Guthrie &amp; Louise Guthrie</b> .....	157
	<i>The University of Sheffield</i>	
10.	Statistical Variations of German Support Verb Constructions in Very Large Corpora	
	<b>Alexander Geyken</b> .....	169
	<i>Berlin-Brandenburg Academy of Sciences and Humanities</i>	
11.	A Case Study in Word Sketches – Czech Verb <i>vidět</i> ‘see’	
	<b>Karel Pala &amp; Pavel Rychlý</b> .....	187
	<i>Masaryk University</i>	
12.	The Lexical Population of Semantic Types in Hanks’s PDEV	
	<b>Silvie Cinková, Martin Holub &amp; Lenka Smejkalová</b> .....	199
	<i>Charles University in Prague</i>	
13.	From Pattern Dictionary to Patternbank	
	<b>Elisabetta Jezek &amp; Francesca Frontini</b> .....	215
	<i>Pavia University</i>	

## **The Lexical Population of Semantic Types in Hanks's PDEV**

---

**Silvie Cinková, Martin Holub &  
Lenka Smejkalová**

### **Abstract**

This contribution reports on an ongoing analysis of the Pattern Dictionary of English Verbs (PDEV, Hanks 2007b) with respect to both its consistency and reproducibility of use by different users. We address, in particular, the assignment of Semantic Type labels to noun collocates of verbs, in a series of experiments conducted at the Institute of Formal and Applied Linguistics of the Charles University in Prague.

### **1. Introduction**

The Pattern Dictionary of English Verbs (PDEV, Hanks & Pustejovsky 2005, Hanks 2007b), implemented by means of Corpus Pattern Analysis (CPA, Hanks 2004, Hanks 2007a), is an ongoing work that aims to map meanings (called 'implicatures' in PDEV) onto patterns of use. The work is still at an early stage, measured by the number of verbs still unprocessed. It has not (so far) been run under the aegis of a large-scale grant, which would provide back-up in the form of a stable team of co-workers. Quite the contrary: PDEV has only been intermittently hosted by other projects. And yet, it has been discovered as an intellectual delight by a number of enthusiasts scattered worldwide. Despite not being a major project with respect to its administrative anchoring, CPA has already exerted considerable impact upon the linguistic community and continues to inspire researchers, who have either started creating their own patterns –

such as Elisabetta Jezek for Italian (Hanks & Jezek 2008, Jezek & Frontini 2010, Jezek & Hanks 2010), and Araceli Alonso and Irene Rigau for Spanish – or have begun to develop tools for a more efficient pattern administration (Horák *et al.* 2006). Harzing's (2010) *Publish or Perish* presents more than 600 topic-relevant references to CPA since its first trackable mention – Hanks (1994).

We regard CPA and its verb patterns as a very promising implementation of Sinclair's inseparable coupling of the paradigmatic with the syntagmatic preferences of word usage in context (Sinclair 1998). To the best of our knowledge, Patrick Hanks's CPA is the only formally encoded lexical resource that explicitly and consistently acknowledges Sinclair's seminal concept of the Lexical Item. We are now taking the first steps towards exploring PDEV with respect to:

- its encoding consistency; and
- whether or not the patterns created are perceived in the same way by different people.

We hope that PDEV will grow into a large-scale project one day and that our findings will be helpful in bringing this about, as well as in keeping the further development consistent.

## 2. The Pattern Dictionary of English Verbs (PDEV)

### 2.1. *Patterns, propositions, implicatures, and collocates*

PDEV consists of three major parts:

- the patterns;
- sample reference data, randomly selected from the BNC, with verb occurrences manually tagged with their respective pattern numbers; and
- a hierarchical inventory of Semantic Types.

Each pattern is formulated as a proposition in which the verb in question is lemmatized and its relevant collocates are classified by means of **semantic labels** or listed as **lexical sets**, depending on whether the respective collocates can be grouped together under the general heading of a **Semantic Type**. Each proposition is paraphrased by a sentence in which the relevant pattern arguments are labelled identically with the proposition part. This paraphrase embodies the **implicature** (or **meaning potential**; see Hanks 1994) activated by that particular pattern.

Each collocate that is not listed as a member of a lexical set is described by a Semantic Type. Semantic Types are sometimes augmented by a **Semantic Role**. For example, the subject of a sentence may have the Semantic Type [[Human]], an intrinsic property of the subject of the verb, but it may also have the Semantic Role «Judge», a property assigned to the Human by the context. The Semantic Types are a finite set of labels hierarchically ordered in what Patrick calls a **shallow semantic ontology** (Hanks & Pustejovsky 2005). For more details see Section 2.3 below. The Semantic Types describe inherent properties of the collocates, such as *Human*, *Artefact*, *Stuff*, *Document*. The Semantic Roles describe properties that are assigned to the word in a particular pattern or context. Only nouns are given Semantic Types.<sup>1</sup>

Verb **complementations** are described according to their surface realizations; for example “to+INF”, “that-clause”, etc.

Therefore, two or more patterns can occur that differ only in the representation of one collocate due to variable surface realizations.

## 2.2. Reference samples

The patterns are created on the basis of a random sample of 250-500 expandable concordance lines, which are then manually sorted according to the corresponding pattern numbers. The randomly selected corpus concordances that contain a verb processed in a particular PDEV entry constitute a **reference sample**. Each occurrence of the verb in the reference sample is manually tagged with a pattern number.

Concordances that fit the patterns smoothly are simply marked with pattern numbers. Inventive uses, irony, and other figurative uses are, whenever possible, associated with a pattern from the selection, but are given the addition “e”, to indicate “exploitation” of a “norm” defined by a pattern. Verb forms (typically participles) that do not have the function of a verb in the concordance (e.g. adjectival and nominal derivatives) are marked “x” (“unmarkable”), while occurrences that do not fit any pattern are marked “u” (“undecidable”).

Whenever possible, an additional set of tagged data is also created by bulk tagging of word sketches (Kilgarriff *et al.* 2004), but this is not part of the reference sample.

All the tagged concordances originate from the BNC50 corpus. The BNC50 corpus is a 50-million-word part of the BNC<sup>2</sup> that was obtained by removing spoken texts and a proportion of written texts. BNC50 contains almost 5,800 verb types occurring in 8 million verb tokens. However, about 41% of all verb tokens represent modal (‘shall,’ ‘can,’ ‘must,’ etc.) or auxiliary (‘will,’ ‘do,’

'have,' and 'be') verbs, which are ignored. The number of lexical verb types in BNC50 is 5,757 and the total number of the corresponding tokens is 4,673,003.

### 2.3. Inventory of Semantic Types

The current version of PDEV contains a shallow semantic ontology with 200 hierarchically ordered Semantic Types, which draws on the Brandeis Semantic Ontology (Pustejovsky *et al.* 2006), but has gradually developed into an independent project. The Semantic Types have arisen from the corpus tagging. The ontology, tailored to the needs of PDEV development, is still subject to revisions, especially regarding the hierarchical order, but the inventory does not undergo any substantial changes during further annotation.

The most essential feature of the PDEV ontology is that it is strictly language-centred. It is designed to represent the linguistic reflection of the world rather than "objective" common knowledge. It is therefore presumably language-specific.

### 3. Consistency checking in PDEV

At the time of writing (April 2010), PDEV comprises 678 completed verb entries with patterns and the corresponding reference sample data. This is 11.8% of all lexical verb types in BNC50. To examine the consistency of the current PDEV version, we selected 30 entries according to their frequency, number of patterns and pattern perplexity (the perplexity was calculated from the differences among the percentages of tagged concordances associated with the respective patterns).<sup>3</sup> We asked two to three annotators (Patrick and ourselves) to tag a new random sample of 50 concordances for each selected verb and we measured the interannotator agreement in order to find out whether anyone else than the author of the patterns is able to distinguish among the patterns.

The overall result of our post-annotation analysis of disagreements in the respective entries was promising: most disagreements observed were accidental errors or oversights, whose number is bound to decrease in proportion to the increasing experience of the annotators. In the second place, the annotators sometimes disagreed on whether a concordance was a **norm** or an **exploitation** of a norm. The pattern numbers usually agreed, which makes this disagreement type less important. Another source of disagreement was the distinction between **markable/unmarkable/undecidable** in participial usages, which bordered on adjectives or nouns, which is also a relatively uninteresting type of disagreement.



Only a few errors pointed at frame inadequacy. Most pattern inadequacies seemed to arise when the annotators intuitively perceived a usage not as an exploitation but as a regular instance of one selected norm (pattern).

The frequency of this last disagreement type suggested that the current patterns would benefit from an update of Semantic Types. To facilitate a possible revision of the ontology, we have performed an analysis of the lexical population of the Semantic Types in the currently tagged PDEV data.

#### 4. Analyzing the lexical population of Semantic Types

##### *4.1. Nouns and Semantic Types in the reference samples*

Each manually tagged verb occurrence refers to a pattern. Each pattern contains one or more collocate slots. The slots open for noun complementations are defined by the Semantic Types, whereas verb complementations, such as content clauses and infinitive and participial clauses are described according to their syntactic features. The verb complementations were not considered in our experiment.

The Semantic Types are present only in the patterns. In other words, they are not explicitly assigned to the collocates in the data. It is therefore not straightforward to find out which nouns populate a particular Semantic Type.

We believe that the explicit linking between the nouns and the Semantic Types is a prerequisite to:

- create definitions for each Semantic Type;
- gain insight about the usefulness of each Semantic Type, as well as about possible unwanted overlaps and gaps in the ontology; and to
- design a well-founded hierarchy of Semantic Types.

Our immediate goal was therefore to obtain a list of nouns and prepositional groups that occurred in the tagged concordances in complementation positions, along with the corresponding Semantic Types mapped automatically from the patterns assigned.

##### *4.2. Parsing the pattern-tagged part of BNC50*

The manually pattern-tagged data comprises approximately 200,000 sentences. This set consists of an average of almost 300 concordances for each of the 678 compiled verbs.

The syntactic description of the respective collocate slots is – more or less consistently – recorded in the patterns. The syntactic description considers positions as subject (which actually includes object in passive clauses), object (i.e. direct object and *to*-object in dative alternating verbs) and adverbial (adverbials and prepositional objects except the dative *to*).

As a start, we parsed the 200,000 pattern-tagged sentences, using the Stanford parser (Klein & Manning 2003, De Marneffe *et al.* 2006) to obtain typed syntactic dependencies. The Stanford parser represents each dependency edge as the syntactic relation between a governing node and its dependent node. In our experiment, the parser was run with the parameter “-outputFormatOptions CCPropagatedDependencies”, and thus the tree structure was not preserved (our parse contained even cycles). See the Addendum for examples.

To improve our procedure, the parsing will be complemented by the Stanford Named-Entity Recognizer (Finkel *et al.* 2005) to make a distinction between e.g. *bush* and *Bush*, since this is naturally relevant for the assignment of the Semantic Types. The Named-Entity Recognizer also makes it possible to treat a multi-word named entity as a single lexical unit.

#### 4.3. Postprocessing

The parsed output was subjected to further postprocessing; the results of which are shown in Table 1.

**Table 1:** Pairs found.

Pairs (SemType, Word) found	223,415
Pairs (SemType, Noun) found	163,807
Pairs (SemType, Pronoun) found	50,063
Unique pairs (SemType, Noun) found	93,366

First, we extracted the nouns dependent on each verb, along with their possible prepositions. Second, the Stanford labels of typed dependencies (edges between the governing and the dependent node in the dependency tree) were converted into the syntactic terms used in PDEV. For instance, the Stanford parser uses the label *nsubj* for the syntactic subject of an active clause and the label *nsubjpass* for the syntactic subject of a passive clause, while the logical subject of a passive clause, which is syntactically rendered as a *by*-object, gets the label *agent*. The labels *nsubj* and *agent* were transformed into the PDEV subject label. What is called adverbial in PDEV, but has the form of a



prepositional phrase, is presented as a prepositional object in the Stanford parser output – see (1):

- (1) Dexy's used to workout together, they abstained from alcohol and drugs.  
 => prep\_from(abstained-9, alcohol-11) prep\_from  
 (abstained-9, drugs-13)

PDEV stores the required preposition directly in a collocate slot, so that the information does not get lost – see (2):

- (2) (for [[**Eventuality** | **Entity**]])<sup>4</sup>

In the next step, we paired the Semantic Types as found in a complementation position (subject, object, or adverbial) of a pattern with the noun found in the parsed text. The output had the following form, (3):

- (3) SemType\_1 | ... | SemType\_n, word

where *n* is typically 1, 2, alternatively 3. Higher numbers hardly ever occur. These pairs represented the basis for creating lists of nouns sorted according to the Semantic Types.

We then merged the pairs into triples, (4):

- (4) (SemType\_1 | ... | SemType\_n, word, frequency)

At this stage, all lines with more than one SemType were ambiguous. These lines reflected cases where the collocate slot explicitly allowed for more than one Semantic Type. This notation is very common in PDEV, as (5) shows, and we had to resolve this ambiguity:

- (5) [[**Human** | **Institution**]] accept [[**Proposition** | **Concept** | **Eventuality**]]

#### 4.4. Resolving ambiguity

We tested four approaches to decide which of the Semantic Types listed within one collocate slot were to be associated with the word extracted from the corpus; see Table 2.

**Table 2:** Assignment Statistics.

Method	1	2	3	4
Unique pairs (SemType, Word) (after disambiguation)	110,715	26,916	110,715	94,394
Unique pairs (SemType, Word) (after smoothing)			3,546	3,905
In %			3.20%	4.14%

## 1. "could be"

We assumed that the word could be a realization of all the Semantic Types listed, and the frequency remained the same in all newly established triples. This approach yielded a high recall but extremely low precision. The sum of frequencies was not equal to the original number of pairs.

## 2. "certainly is"

When we did not consider the ambiguous triples, we obtained high precision but extremely low recall and the sum of frequencies was not equal to the original number of pairs.

## 3. "equal chance"

In the third test we divided the frequency between the competing Semantic Types evenly. The sum of frequencies matched the original number of pairs.

## 4. "combined procedure"

From the lists generated by the "certainly is" approach we obtained the a priori likelihoods of the association of a given word with different Semantic Types. We divided the total frequency according to their ratio. This approach was theoretically the soundest one, but it was biased by the fact that the Semantic Types that had never occurred alone in the collocate slots but only in groups (typically: Human and Institution) were systematically getting lower ratings than Semantic Types that had occurred alone with the same word in different patterns. We introduced a correction by the equal division of frequencies like in the "equal chance" approach. The sum of frequencies then matched the original number of pairs.

Given the results from 3 and 4, we can calculate the pointwise mutual information (PMI) for each pair (SemType, Word). PMI is defined by the following formula:

$$\text{PMI}(\text{SemType}, \text{Word}) = \log_2 ( p(\text{SemType}, \text{Word}) / (p(\text{SemType}) * p(\text{Word})) )$$



PMI enhances the simple frequency counts with the information on which words are strongly related to a particular Semantic Type. Then the results can be filtered by removing both rare pairs and pairs with negative or very low PMI. With the frequency limit at 5 and the minimum PMI at 0.5 the entire output list shrinks to about 4% of the original size. See for an example Table 3.

The resulting lists can be used to determine the words most likely associated with a particular Semantic Type as well as to find the most likely Semantic Type associated with a word.

**Table 3:** Assignment of Semantic Types to the word *car*.

	SemType	Method:	1	2	3	4	PMI(3)	PMI(4)			
<i>car</i>	Human		52	10	28.17	45.31	-0.67	-0.52	FALSE	FALSE	FALSE
<i>car</i>	Artefact		34	24	26.92	33	4.47	4.96	TRUE	TRUE	TRUE
<i>car</i>	Physical Object		27	15	20	24.29	2.54	2.64	TRUE	TRUE	TRUE
<i>car</i>	Road Vehicle		27	1	14	27	7.84	8.67	TRUE	TRUE	TRUE
<i>car</i>	Vehicle		32	0	12.83	1	4.56	2.23	TRUE	FALSE	FALSE
<i>car</i>	Plane		24	0	12	0	5.95	-inf	TRUE	FALSE	FALSE
<i>car</i>	Anything		6	5	5.5	5.33	-0.95	-1.01	FALSE	FALSE	FALSE
<i>car</i>	Entity		6	4	4.75	4.86	-0.32	-0.37	FALSE	FALSE	FALSE
<i>car</i>	Eventuality		5	3	4	3.66	-1.21	-1.39	FALSE	FALSE	FALSE
<i>car</i>	Activity		8	1	3.42	1.38	-0.54	-1.95	FALSE	FALSE	FALSE
<i>car</i>	Animate		9	0	3.17	0.5	2.46	0.69	FALSE	FALSE	FALSE
<i>car</i>	Location		4	1	2.17	1.15	-1.05	-1.91	FALSE	FALSE	FALSE
<i>car</i>	Stuff		6	0	2.17	0	1.07	-inf	FALSE	FALSE	FALSE
<i>car</i>	Building		6	0	1.5	0	3.04	-inf	FALSE	FALSE	FALSE
<i>car</i>	Institution		3	0	1.5	0	-3.89	-inf	FALSE	FALSE	FALSE
<i>car</i>	Room		6	0	1.5	0	5.28	-inf	FALSE	FALSE	FALSE
<i>car</i>	Concept		2	1	1.25	1.03	-0.18	-0.45	FALSE	FALSE	FALSE
<i>car</i>	Event		3	0	1.08	0.5	-1.62	-2.76	FALSE	FALSE	FALSE
<i>car</i>	Action		2	0	1	0.5	-1.83	-3.16	FALSE	FALSE	FALSE
<i>car</i>	Boat		2	0	1	0	2.68	-inf	FALSE	FALSE	FALSE
<i>car</i>	Dog		1	1	1	1	3.74	4.23	FALSE	FALSE	FALSE
<i>car</i>	Firearm		1	1	1	1	3.22	3.25	FALSE	FALSE	FALSE
<i>car</i>	Animal		3	0	0.92	0	-1.16	-inf	FALSE	FALSE	FALSE
<i>car</i>	Human Group		2	0	0.83	0.5	-1.11	-1.28	FALSE	FALSE	FALSE
<i>car</i>	Food		1	0	0.5	0.5	1.24	1.23	FALSE	FALSE	FALSE
<i>car</i>	Ship		1	0	0.5	0.5	3.76	4.22	FALSE	FALSE	FALSE
<i>car</i>	State		1	0	0.5	0.5	-2.1	-1.51	FALSE	FALSE	FALSE
<i>car</i>	Vehicle Group		1	0	0.5	0.5	4.41	4.85	FALSE	FALSE	FALSE
<i>car</i>	Noise		1	0	0.33	0	3.47	-inf	FALSE	FALSE	FALSE

## 5. Conclusion

At the moment, there is no other way to assess the credibility of the resulting lists but visual control, since there is no manually annotated testing data available. It is worth considering creating such data by assigning Semantic Types to collocates directly in the reference sample data when assigning pattern numbers to verb tokens. Annotating collocates with Semantic Types would also enable closer analysis of the norm-exploitation annotator disagreements, as it is mostly the presence of an unexpected Semantic Type that decides.

Our pilot analysis of the lexical population of Semantic Types, which we performed as part of a broader analysis of the patterns and the reference sample data in PDEV, was intended to facilitate a future large-scale expansion of PDEV, as well as to make sure that PDEV will be usable as a standard lexical resource in automatic natural language processing.

### Acknowledgements

This project was kindly supported by EuroMatrixPlus (FP7-ICT-2007-3-231720 of the EU and 7E09003 of the Czech Republic), the Czech Science Foundation GA P406/2010/0875, and the Czech Ministry of Education MSM 0021620838.

### Notes

<sup>1</sup> In fact, also anaphorical expressions with noun antecedents/postcedents are assigned a Semantic Type, according to their antecedent/postcedent noun.

<sup>2</sup> See <http://www.natcorp.ox.ac.uk/>.

<sup>3</sup> The sample data for a few entries was taken from the Wall Street Journal-Penn Treebank Corpus (Marcus *et al.* 1993), to assess whether or not the BNC50-based patterns are corpus-specific.

<sup>4</sup> This collocate is then classified as adverbial.

### References

- De Marneffe, M.-C., B. MacCartney & C. D. Manning.** 2006. 'Generating Typed Dependency Parses from Phrase Structure Parses' in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC '06)*. Genoa: LREC, 449–454.
- Finkel, J. R., T. Grenager & C. D. Manning.** 2005. 'Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling' in *Proceedings of the 43<sup>rd</sup> Annual Meeting of the Association for Computational Linguistics*. Ann Arbor, MI: ACL, 363–370.
- Hanks, P.** 1994. 'Linguistic Norms and Pragmatic Exploitations, or Why Lexicographers need Prototype Theory, and Vice Versa' in F. Kiefer, G. Kiss & J. Pajzs (eds.). 1994. *Papers in Computational Lexicography: COMPLEX '94*. Budapest: Linguistics Institute, Hungarian Academy of Sciences, 89–113.
- Hanks, P.** 2004. 'Corpus Pattern Analysis' in G. Williams & S. Vessier (eds.). 2004, 87–97.



- Hanks, P.** 2007a. *Corpus Pattern Analysis (CPA) – Project Page*. Online at <http://nlp.fi.muni.cz/projects/cpa/>.
- Hanks, P.** 2007b. *Pattern Dictionary of English Verbs (PDEV) – Project Page*. Online at <http://deb.fi.muni.cz/pdev/>.
- Hanks, P. & E. Jezek.** 2008. 'Shimmering Lexical Sets' in E. Bernal & J. De-Cesaris (eds.). 2008. *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)* (Sèrie Activitats 20). Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 391–402.
- Hanks, P. & J. Pustejovsky.** 2005. 'A Pattern Dictionary for Natural Language Processing'. *Revue française de linguistique appliquée* 10.2: 63–82.
- Harzing, A. W.** 2010. *Publish or Perish*. Online at <http://www.harzing.com/pop.htm>.
- Horák, A., K. Pala, A. Rambousek & P. Rychlý.** 2006. 'New Clients for Dictionary Writing on the DEB Platform' in G.-M. de Schryver (ed.). 2006. *DWS 2006: Proceedings of the Fourth International Workshop on Dictionary Writing Systems, Tuesday 5th September 2006, Turin, Italy (Pre-EURALEX 2006)*. Pretoria: (SF)<sup>2</sup> Press, 41–46.
- Jezek, E. & F. Frontini.** 2010. 'From Pattern Dictionary to Patternbank' in G.-M. de Schryver (ed.). 2010. *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*. Kampala: Menha Publishers, 215–239.
- Jezek, E. & P. Hanks.** 2010. 'What Lexical Sets Tell Us about Conceptual Categories'. *Lexis: E-Journal in English Lexicology* 4 (Corpus Linguistics and the Lexicon / La linguistique de corpus et le lexique), 7–22. Online at <http://screcherche.univ-lyon3.fr/lexis/spip.php?article136>.
- Kilgarrieff, A., P. Rychlý, P. Smrž & D. Tugwell.** 2004. 'The Sketch Engine' in G. Williams & S. Vessier (eds.). 2004, 105–116.
- Klein, D. & C. D. Manning.** 2003. 'Fast Exact Inference with a Factored Model for Natural Language Parsing' in *Advances in Neural Information Processing Systems 15 (NIPS 2002)*. Cambridge, MA: MIT Press, 3–10.
- Marcus, M. P., B. Santorini & M. A. Marcinkiewicz.** 1993. 'Building a large annotated corpus of English: The Penn Treebank'. *Computational Linguistics* 19.2: 313–330.
- Pustejovsky, J., C. Havasi, J. Littman, A. Rumshisky & M. Verhagen.** 2006. 'Towards a Generative Lexical Resource: The Brandeis Semantic Ontology' in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC '06)*. Genoa: LREC, 1702–1705.
- Sinclair, J. M.** 1998. 'The Lexical Item' in E. Weigand (ed.). 1998. *Contrastive Lexical Semantics*. Amsterdam: John Benjamins, 1–24.

**Williams, G. & S. Vessier** (eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004*. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

## Addendum

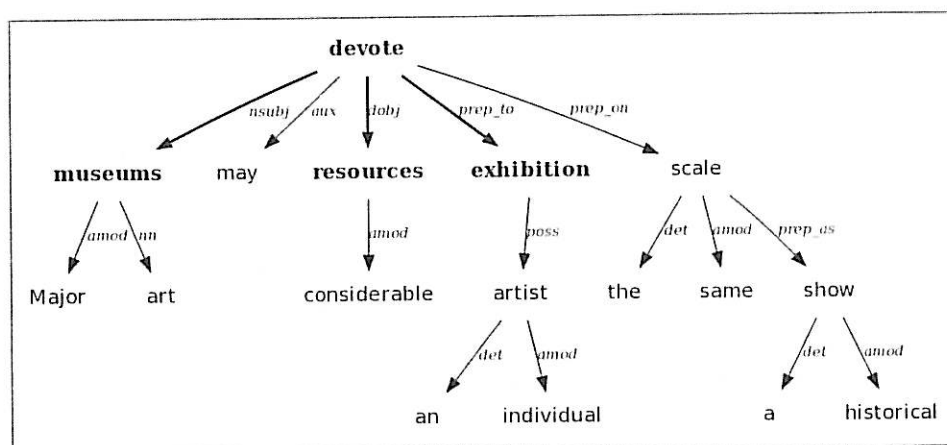
1.

verb: **devote**, pattern: 1 (46%)

pattern\_def: **[[Human | Institution]] devote [[Resource]] {to [[Activity]]}**

implicature: **[[Human | Institution]] expend [[Resource]] on [[Activity]]**

**Example 1:** Major art **museums** may devote considerable **resources** to an individual artist's **exhibition**, on the same scale as a historical show.



**Figure 1:** Typed-dependencies output of the Stanford parser for Example 1.

### Stanford typed-dependencies:

amod(museums-3, Major-1)  
 nn(museums-3, art-2)  
**nsubj(devote-5, museums-3)**  
 aux(devote-5, may-4)  
 amod(resources-7, considerable-6)  
**dobj(devote-5, resources-7)**  
 det(artist-11, an-9)

amod(artist-11, individual-10)  
 poss(exhibition-13, artist-11)  
**prep\_to(devote-5, exhibition-13)**  
 det(scale-18, the-16)  
 amod(scale-18, same-17)  
**prep\_on(devote-5, scale-18)**  
 det(show-22, a-20)  
 amod(show-22, historical-21)  
 prep\_as(scale-18, show-22)

PDEV typed-dependencies:

subject: museums/NNS  
 object: resources/NNS  
 adverbial: exhibition/NN rel: prep\_to  
 adverbial: scale/NN rel: prep\_on

Mapping output:

Human|Institution - museums/NNS  
 Resource - resources/NNS  
 Activity - exhibition/NN

2.

verb: **devote**, pattern: 2 (18%)  
 pattern\_def: [[{**Human** = **Author**} | {**Institution** = **Newspaper** | **Journal**}]  
 devote [[**Document Part** | **Document**]] {to [[**Anything** =  
**Topic**]]}  
 implicature: [[{**Human** = **Author**} | {**Institution** = **Newspaper** | **Journal**} |  
**Document**]] gives priority to discussion or elaboration of [[**Any-**  
**thing** = **Topic**]] in [[**Document Part** | **Document**]]

**Example 2:** In a month when Harpers & Queen **magazine** devotes several **pages** to the lost **art** of manners, it may seem strange that parents should send their daughters from the most polite nation in the world to this newly graceless country.



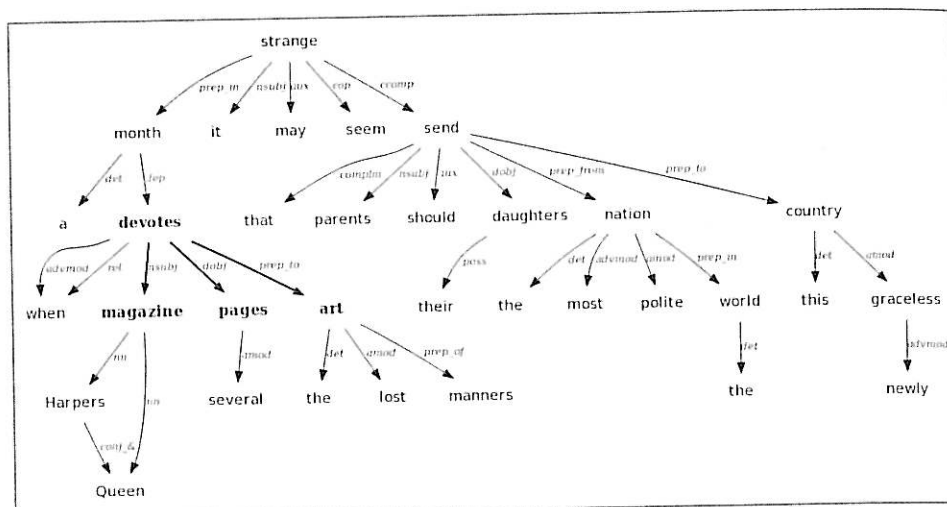


Figure 2: Typed-dependencies output of the Stanford parser for Example 2.

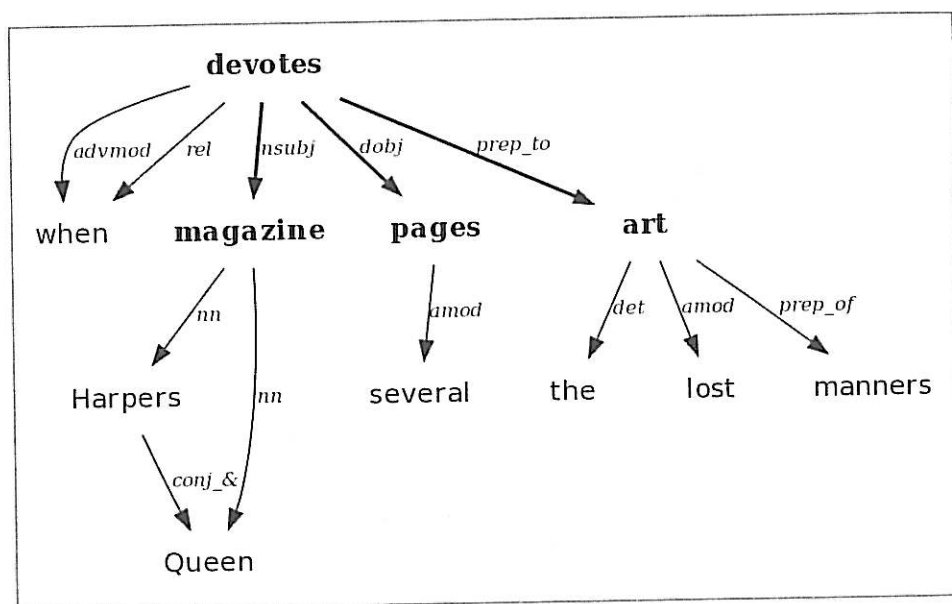


Figure 3: Typed-dependencies output of the Stanford parser for Example 2 – detail.

#### Stanford typed-dependencies:

det(month-3, a-2)  
 prep\_in(strange-22, month-3)

**advmod(devotes-9, when-4)**

rel(devotes-9, when-4)

nn(magazine-8, Harpers-5)

conj\_&amp;(Harpers-5, Queen-7)

nn(magazine-8, Queen-7)

**nsubj(devotes-9, magazine-8)**

dep(month-3, devotes-9)

amod(pages-11, several-10)

**dobj(devotes-9, pages-11)**

det(art-15, the-13)

amod(art-15, lost-14)

**prep\_to(devotes-9, art-15)**

prep\_of(art-15, manners-17)

nsubj(strange-22, it-19)

aux(strange-22, may-20)

cop(strange-22, seem-21)

complm(send-26, that-23)

nsubj(send-26, parents-24)

aux(send-26, should-25)

ccomp(strange-22, send-26)

poss(daughters-28, their-27)

dobj(send-26, daughters-28)

det(nation-33, the-30)

advmod(nation-33, most-31)

amod(nation-33, polite-32)

prep\_from(send-26, nation-33)

det(world-36, the-35)

prep\_in(nation-33, world-36)

det(country-41, this-38)

advmod(graceless-40, newly-39)

amod(country-41, graceless-40)

prep\_to(send-26, country-41)

PDEV typed-dependencies:

adverbial: when/WRB rel: advmod

subject: magazine/NN

object: pages/NNS

adverbial: art/NN rel: prep\_to

Mapping output:

Human|Institution - magazine/NN

Document Part|Document - pages/NNS

Anything – art/NN