PDEV-lemon: a Linked Data implementation of the Pattern Dictionary of English Verbs based on the Lemon model

Ismaïl El Maarouf, Jane Bradbury, Patrick Hanks

University of Wolverhampton
i.el-maarouf@wlv.ac.uk, J.Bradbury3@wlv.ac.uk, patrick.w.hanks@gmail.com

Abstract

PDEV-Lemon is the Linguistic Linked Data resource built from PDEV (Pattern Dictionary of English Verbs), using the Lemon lexicon model (Mccrae et al., 2012). PDEV is a dictionary which provides insight into how verbs collocate with nouns and words using an empirically well-founded apparatus of syntactic and semantic categories. It is a valuable resource for Natural Language Processing because it specifies in detail the contextual conditions that determine the meaning of a word. Over 1000 verbs have been analysed to date. PDEV-Lemon is built using the Lemon model, the LExicon Model for ONtologies.

Keywords: Semantic Web, Linguistic Linked Data, PDEV, CPA, lemon model, Lexicography, Lexicon, Ontology

Introduction

This paper introduces the first Semantic Web adaptation of PDEV (Pattern Dictionary of English Verbs; `http://pdev.org.uk`), using the Lemon lexicon model (Mccrae et al., 2012). PDEV is a dictionary which provides insight into how verbs collocate with nouns and other words using an empirically well-founded apparatus of syntactic and semantic categories. PDEV is a valuable resource for NLP because it specifies in detail the contextual conditions that determine meaning of a word. Thus the main motivation for building a Semantic-Web-compliant resource is to provide the Semantic Web and the NLP communities with an easier access to PDEV.

Section 1 provides an overview of PDEV, and of its specific characteristics. Section 2 describes PDEV-lemon. Section 3 reviews applications of PDEV-Lemon.

1. Background for the Pattern Dictionary of English Verbs

PDEV is an electronic resource (work in progress) consisting of an inventory of English verbs that identifies all the *normal patterns of use* of each verb and associates each pattern with a meaning, implicature, and/or entailment. Patterns are the conventional collocational structures that people use, and are identified from large corpora using a technique named CPA (Corpus Pattern Analysis (Hanks, 2013)).

This technique provides means for identifying patterns and for mapping meaning onto words in text. It is based on the Theory of Norms and Exploitations (TNE) (Hanks, 2004; Hanks, 2013; Hanks and Pustejovsky, 2005). TNE proposes to classify word uses in two main categories: *norms*, which are conventional uses of a word, frequently observed, and *exploitations*, which consist in deviations from those norms. This double helix theory of language use transcends, and sheds a new light on, traditional phenomena such as idioms, metaphors, or coercions.

TNE is a theory of language which emerged from extensive corpus analysis in corpus linguistics, but it is also influenced by the theory of the Generative Lexicon (Pustejovsky, 1995), Wilks's theory of Preference Semantics (Wilks, 1975), and Frame Semantics (Fillmore, 1985; Baker et al., 1998). Its roots in corpus linguistics are set in Sinclair's ground-breaking work on corpus analysis and collocations (Sinclair, 2010; Sinclair, 1991; Sinclair, 2004), the COBUILD project (Sinclair, 1987), and the Hector project (Atkins, 1993; Hanks, 1994). TNE therefore bridges the gap between corpus linguistics and semantic theories of the lexicon by combining insights from both perspectives to provide a model for empirically well-founded lexical resources.

PDEV, which draws on TNE, offers the analysis of over 1,000 verbs, over 4,000 patterns, and more than 100,000 annotated lines tagged as evidence, out of the list of 5793 verbs retained in this lexicon. PDEV lexicographers identify patterns using large corpora such as the British National Corpus[1] (BNC) and the Oxford English Corpus[2] (OEC).

The analysis is corpus-driven and is therefore empirically well-founded. Noun collocates are organized into lexical sets in relation to the verbs with which they habitually co-occur. The lexical sets are organized according to semantic type, in a shallow ontology. The shallow ontology consists of fewer than 250 semantic types, which, it turns out, are sufficient to disambiguate the senses of the verbs analysed so far. The ontology is hierarchically organized, with *Entity* and *Eventuality* being the two top types.

2. PDEV-Lemon

2.1. Specifications

PDEV is being developed over a dedicated system which includes Javascript user interfaces, an SQL database, the DEB platform[3], and the SketchEngine corpus query system (Kilgarriff and Rychly, 2010). PDEV-Lemon is a derived resource in RDF based on the Lemon lexicon model. The development of a Semantic Web compliant resource was based on the following specifications:

[1] `http://www.natcorp.ox.ac.uk/`

[2] `http://www.oxforddictionaries.com/words/the-oxford-english-corpus`

[3] `http://deb.fi.muni.cz/index.php`

- Simplicity and usability of the lexicon.
- Faithfulness to the principles of PDEV.
- Identification of unstructured data in the lexicon.
- Adaptability allowing future automatic generation of dumps.
- Availability to the NLP community at large.

To create the resource, the following main tasks were carried out:

1. Study of the Lemon model and adaptation to the PDEV lexicon.
2. Creation of an ontology framework to provide external vocabulary for PDEV specific descriptive categories.
3. Creation of a system to process the data and generate RDF dumps of the database on demand.

2.2. Lemon: the backbone

One of the standard models for building Semantic Web machine-readable dictionaries is Lemon (The Lexicon Model for Ontologies (Mccrae et al., 2012; Buitelaar et al., 2011))[4]. This RDF-native model provides the general structure and features to enable an easy instantiation of a lexicon using an ontology framework such as OWL[5]. In Lemon, it is possible to create word entries, provide lexical variants, specify word morphology parts, syntactic frames, lexical meaning, and much more.

The primary issue when using Lemon to model syntactic resources is how to map a syntactic frame, selected by a lexical entry, to a meaning. Lemon does not provide direct links between a frame and a lexical sense: a lexical sense combines a lexical entry with a reference to a meaning (defined in an external ontology), and the contextual environment in which this meaning occurs is underspecified. The only way to map frames to a lexical meaning in Lemon is by the mediation of the frame's arguments, provided correlated units exist in external ontologies: the lexical sense is in this case indirectly induced [6].

This is not satisfactory from the point of view of understanding and processing the meaning of texts. One of PDEV's contribution is to show the ways in which syntax and semantics do not map neatly onto each other, because of phraseology. As an example, an idiom cannot be mapped because its parts are not interpreted as concepts: the whole meaning of an idiom is different from the sum of the meaning of its parts. Since a large part of language is phraseological, it seems safer to alter the Lemon model to allow for a direct mapping of frame with lexical sense.

We therefore provide two new Object properties to link one *Frame* to one *Lexical Sense*. Following *lemon:senseOf* and *lemon:sense*, which are used to map lexical senses with lexical entries, PDEV-Lemon adds *:frameSense* and *:isFrameSenseOf* to map lexical senses with frames (Fig. 1).

[4] http://lemon-model.net

[5] http://www.w3.org/TR/owl-features/

[6] for examples of implementation see http://lemon-model.net/lexica/lexica.php

```
:frameSense
  rdf:type rdf:Property,owl:ObjectProperty ;
  rdfs:label "Frame Sense"@en ;
  rdfs:comment "Links to the lexical sense of a
frame"@en ;
  rdfs:domain :Frame ;
  rdfs:range :LexicalSense .

:isFrameSenseOf
  rdf:type rdf:Property,owl:ObjectProperty ;
  rdfs:label "Frame Sense of"@en ;
  rdfs:comment "Indicate that a sense is realised
by the given frame"@en ;
  rdfs:domain :LexicalSense ;
  rdfs:range :Frame ;
  owl:inverseOf :frameSense .
```

Figure 1: *frameSense* and *isFrameSenseOf* properties

2.3. The PDEV-Lemon entry

PDEV-Lemon makes use of the Lemon core to create the lexicon and its individuals. As an example, Fig. 2 shows the PDEV-Lemon entry, in TURTLE syntax, for *organize*.

```
pdevl:PDEV_LexicalEntry_organize
  rdf:type lemon:LexicalEntry, owl:NamedIndividual;
  lexinfo:partOfSpeech lexinfo:verb ;
  rdfs:label "PDEV Lexical Entry organize"@eng ;
  lemon:canonicalForm
    pdevl:PDEV_LexicalEntry_organize_CanonicalForm ;
  ps:lexicalFrequencyOf
    pdevl:PDEV_LexicalEntry_organize_sampleSize ;
  ps:lexicalFrequencyOf
    pdevl:PDEV_LexicalEntry_organize_bncFreq ;
  ps:lexicalFrequencyOf
    pdevl:PDEV_LexicalEntry_organize_bnc50Freq ;
  ps:lexicalFrequencyOf
    pdevl:PDEV_LexicalEntry_organize_oecFreq ;
  lemon:sense pdevl:PDEV_Implicature_organize_1;
  lemon:sense pdevl:PDEV_Implicature_organize_2;
  lemon:sense pdevl:PDEV_Implicature_organize_3;
  lemon:sense pdevl:PDEV_Implicature_organize_4;
  lemon:sense pdevl:PDEV_Implicature_organize_5;
  lemon:sense pdevl:PDEV_Implicature_organize_6;
  lemon:synBehavior pdevl:PDEV_Pattern_organize_1;
  lemon:synBehavior pdevl:PDEV_Pattern_organize_2;
  lemon:synBehavior pdevl:PDEV_Pattern_organize_3;
  lemon:synBehavior pdevl:PDEV_Pattern_organize_4;
  lemon:synBehavior pdevl:PDEV_Pattern_organize_5;
  lemon:synBehavior pdevl:PDEV_Pattern_organize_6;
  lemon:language "eng".
```

Figure 2: Example of lexical entry: the example of *organize* *(ps stands for pdevl-structure ontology.)*

As can be seen, an entry contains information regarding part of speech, language, canonical form, lexical frequency in a given corpus, links to lexical senses and to syntactic frames. Most properties link the lexical entry to URIs which describe lexical information, such as the form (Fig. 3), which itself may contain a link to a variant form.

2.4. The PDEV-Lemon pattern

A PDEV entry contains at least one pattern, from the *lemon:Frame* class. A PDEV pattern is an abstract syntactic and semantic representation of a word's context, and is rephrased as an implicature. More specifically, what PDEV offers through the pattern structure is a set of collocational preferences mapped onto syntactic arguments which are interconnected through the pattern. The structure of a pattern, as well as the properties and categories of its arguments,

```
pdevl:PDEV_LexicalEntry_organize_CanonicalForm
  rdf:type lemon:Form, owl:NamedIndividual;
  lemon:writtenRep "organize"@eng ;
  lemon:formVariant
    pdevl:PDEV_LexicalEntry_organise_VariantForm;
  rdfs:label "Canonical Form of organize"@eng .
pdevl:PDEV_LexicalEntry_organise_VariantForm
  rdf:type lemon:Form, owl:NamedIndividual;
  lemon:writtenRep "organise"@eng ;
  rdfs:label "Variant Form of organize"@eng;
  ps:formVariantof
    pdevl:PDEV_LexicalEntry_organize_CanonicalForm.
```

Figure 3: Instance of the *Form* class: *organize*
(ps stands for pdevl-structure ontology)

have been developed on the basis of observation of empirical data to meet the needs of lexicographers for modeling a pattern's contextual features appropriately.

The structure of a verb pattern is based on the SPOCA model from Systemic Functional Grammar (Halliday, 1994). A verb pattern may consist of arguments from any of the following clause roles: a Subject, a Predicator, an Object, a Complement, and an Adverbial (as well as Indirect objects).

Each argument may be structured into components such as an introductory word (like a preposition), a specifier (like a determiner) and a head. Each component can in turn be represented according to several layers:

- a Grammatical Category (noun, *-ing* forms or quotes),
- a Semantic Type (ST; Human, Animal),
- a Contextual Role (CR; Judge, Plaintiff),
- a Lexical Set (LS; instances of lexical words).

Some of the grammatical categories have been based on the LexInfo ontology (Cimiano et al., 2011), where relevant. In some cases (mainly for adverbials), there can be more than one obligatory argument for the same clause role; ST, CR and LS can also express alternative realizations.

All PDEV patterns are connected to a set of tagged concordances taken from a random sample of the BNC (usually 250 lines) and accessible online on the PDEV public access (`http://pdev.org.uk`). There are two types of links: normal uses of the pattern and exploitations. These links between the dictionary and the BNC have been preserved in PDEV-Lemon. In addition, frequency information for each pattern's normal use and exploitation have been added to PDEV-Lemon. For each entry, it is therefore possible to produce percentages for each pattern.

Fig. 4 gives the full representation in TURTLE syntax of pattern 2 of the verb *zap*.

2.5. The PDEV-Lemon Linked Data suite

An instance of a pattern is always linked to an instance of the *lemon:LexicalSense* class, the reference of which is a unique concept in an external ontology, named *pdev-lemon-PatSenses*. This ontology describes the senses referred to by Frames using their implicatures. Senses have also been

```
pdevl:PDEV_Arg_S_M_zap_2
  rdf:type lemon:Argument, owl:NamedIndividual ;
  ps:syntacticCategory ps:NounPhrase ;
  ps:SemanticType po:PdevSemanticType_36 ;
  ps:argStatus ps:Prototypical .
pdevl:PDEV_Arg_A_M-1_zap_2
  rdf:type lemon:Argument, owl:NamedIndividual ;
  ps:preposition pdevl:PDEP_LexicalEntry_10 ;
  ps:syntacticCategory ps:PrepositionalPhrase ;
  ps:SemanticType po:PdevSemanticType_11 ;
  ps:ContextualRole pt:PdevContextualRole_623 ;
  ps:argStatus ps:Prototypical .
pdevl:PDEV_Arg_A_M-2_zap_2
  rdf:type lemon:Argument, owl:NamedIndividual ;
  ps:preposition pdevl:PDEP_LexicalEntry_27 ;
  ps:syntacticCategory ps:PrepositionalPhrase ;
  ps:SemanticType po:PdevSemanticType_11 ;
  ps:ContextualRole pt:PdevContextualRole_624 ;
  ps:argStatus ps:Prototypical .
pdevl:PDEV_Pattern_zap_2
  rdf:type lemon:Frame, owl:NamedIndividual ;
  ps:isNoObj "true" ;
  ps:senseFrequencyOf pdevl:Freq_norm_zap_2 ;
  ps:senseFrequencyOf pdevl:Freq_exploitation_zap_2 ;
  ps:subject pdevl:PDEV_Arg_S_M_zap_2;
  ps:Predicator "zap" ;
  ps:adverbial pdevl:PDEV_Arg_A_M-1_zap_2;
  ps:adverbial pdevl:PDEV_Arg_A_M-2_zap_2;
  pdevl:frameSense pdevl:PDEV_LexicalSense_zap_2 .
pdevl:Freq_norm_zap_2
  rdf:type ps:Frequency, owl:NamedIndividual ;
  ps:ofCorpus "BNC50" ;
  ps:frequencyValue 2 .
pdevl:Freq_exploitation_zap_2
  rdf:type ps:Frequency, owl:NamedIndividual ;
  ps:ofCorpus "BNC50" ;
  ps:frequencyValue 0 .
```

Figure 4: Example of a PDEV-Lemon pattern: *zap*
(ps, po, pt are prefixes which stand for pdevl-structure, pdevl-CPASO, pdevl-CoRoTaxo ontologies, respectively.)

grouped into semantic classes (530 patterns have been classified), and linked to external resources such as FrameNet[7] (1492 links manually identified by lexicographers). In addition, the PDEV-Lemon resource includes four ontologies which list the descriptive categories used to characterize patterns and entries.

- *pdev-lemon-domain*: describes the domains used to characterize PDEV patterns;
- *pdev-lemon-register*: describes the registers used to characterize PDEV patterns;
- *pdev-lemon-CPASO*: describes the Semantic Types used to characterize PDEV patterns;
- *pdev-lemon-CoRoTaxo*: describes the taxonomy of Contextual Roles used to characterize PDEV patterns.

Finally, *pdev-lemon-structure* specifies the OWL classes and properties needed to alter the Lemon model (23 classes, 22 Object properties, and 8 Datatype properties). The resource *pdev-lemon* contains the dictionary information. All seven developed resources are also available as linked data from `http://pdev.org.uk/PDEVLEMON.html`. Table 1 lists the most frequently used properties in the whole resource.

PDEV-Lemon consists of 217,634 triples, 3702 patterns and 10799 arguments. It contains lexical entries for 984

[7] `http://framenet.icsi.berkeley.edu/`

Frequency	Property
80956	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
11309	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/argStatus>
11298	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/frequencyValue>
11298	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/ofCorpus>
11234	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/syntacticCategory>
9388	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/SemanticType>
7403	<http://www.monnet-project.eu/lemon#value>
7402	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/senseFrequencyOf>
7402	<http://www.monnet-project.eu/lemon#example>
6959	<http://www.w3.org/2000/01/rdf-schema#label>
5301	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/subject>
3896	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/lexicalFrequencyOf>
3702	<http://www.monnet-project.eu/lemon#sense>
3702	<http://www.monnet-project.eu/lemon#synBehavior>
3701	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/frameSense>
3701	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/isFrameSenseOf>
3701	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/Predicator>
3701	<http://www.monnet-project.eu/lemon#reference>
3363	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/directObject>
2606	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/adverbial>
2317	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/preposition>
2077	<http://lexinfo.net/ontology/2.0/lexinfo#partOfSpeech>
2077	<http://www.monnet-project.eu/lemon#entry>
1676	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/ContextualRole>
1176	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/isNoObj>
1106	<http://www.monnet-project.eu/lemon#language>
986	<http://www.monnet-project.eu/lemon#writtenRep>
974	<http://www.monnet-project.eu/lemon#canonicalForm>
973	<http://pdev.org.uk/pdevlemon/pdevlemon-structure/LexicalSet>
908	<http://www.monnet-project.eu/lemon#optional>

Table 1: Most frequent properties used in PDEV-Lemon

verbs, 1030 nouns, and 93 prepositions. It contains 94 domains, 34 registers, 248 semantic types, 662 contextual roles, and 3702 pattern senses.

3. Applications of PDEV-Lemon

As a Linguistic Linked Data resource, PDEV-Lemon will enable the NLP community as well as the Semantic Web community to extract pattern information and integrate it in various applications and resources.

3.1. Leveraging resources

PDEV patterns include nouns and prepositions in argument slots, which have been turned into lexical entries in PDEV-Lemon. This is a first step on which to bootstrap further analyses of noun and preposition pattern dictionaries (Litkowski, 2012).

PDEV is not the only project based on TNE. Italian (Jezek and Frontini, 2010) and Spanish (Renau and Battaner, 2012) versions have been developed with identical resources and tools. The Italian pattern dictionary contains about 3000 patterns for more than 800 verbs and a Spanish version has been developed on more than 150 verbs. PDEV-Lemon has therefore a potential to connect these languages, given that they use the same descriptive apparatus. Such a multiligual resource could be an important asset for research in Machine Translation.

From a more general perspective, PDEV-Lemon allows to connect more easily other lexical resources such as those developed in the UBY framework[8], particularly FrameNet and VerbNet. Immediate plans include the use of FrameNet links (from the *pdev-lemon-PatSenses* ontology) manually defined by lexicographers to leverage information from the Framenet resource and FrameNet annotated corpora. Since FrameNet frames are matched with PDEV patterns, FrameNet can also benefit from an accurate description of the context where lexical units trigger frames. Beyond that, it is also possible to imagine to connect PDEV-Lemon to the resources to which FrameNet is connected, e.g. Wordnet, Verbnet.

3.2. Applications

Pattern discovery and disambiguation.

PDEV patterns involve both an analysis of a word's context and its correlation with meaning. The main goal of the

[8] http://www.ukp.tu-darmstadt.de/data/lexical-resources/uby/

DVC project[9] (Disambiguating Verbs by Collocation) is to build a pattern dictionary of 3,000 verbs following the principles of TNE. Since PDEV-Lemon can be automatically generated, the results of the DVC project will be made regularly available through its future releases.

One of the motivations for building patterns is the observation that while words in isolation are highly ambiguous, corpus analysis shows that patterns are mostly unambiguous. PDEV therefore tackles the Word Sense Disambiguation (WSD) (Navigli, 2009) problem by sidestepping it: instead of asking about the meaning of words, it asks about the meaning of the patterns in which words are used. This resource therefore provides the opportunity to develop new semantic parsers, which can identify patterns in texts as well as their arugments. Preliminary research performed on automatic pattern disambiguation provides promising results (El Maarouf et al., 2014).

Making PDEV-Lemon widely available will be a means to allow researchers to test these patterns at a large scale.

Modeling links between corpus and lexical resources.

PDEV-Lemon includes links to examples of norms and exploitations for each pattern (more than 100,000 concordance lines). In PDEV-Lemon, these links refer to whole concordances available on the PDEV public access. Future plans include specifying the structure of these concordances and map PDEV pattern arguments onto tokens. This will be achieved by taking avantage of standards for Linked annotated resources (such as based on the NLP Interchange Format model[10]) and PDEV-Lemon.

Language Learning.

Beyond NLP applications, PDEV can also be used in pedagogical applications such as tools and interfaces to improve learners' command of idiomaticity, to design a syllabus, and for error correction. For example, the detailed mapping of how certain Semantic Types and adverbial patterns are preferred in certain patterns of certain verbs can help L2 (non-native speakers) students to achieve a high level of naturalness in their speech and writing. A resource such as PDEV-Lemon will facilitate the development of tools for this community.

4. Conclusion

This paper has presented PDEV-Lemon, a new Linguistic Linked Data resource based on PDEV. PDEV is a dictionary of English verbs that identifies all the normal patterns of use of each verb. Patterns are conventional collocational structures linked to a unique meaning, as identified from large corpora.

PDEV-Lemon comes with a suite of OWL ontologies which characterize descriptive categories used in PDEV patterns: domains, registers, semantic types, contextual roles, and pattern senses.

PDEV-Lemon was developed to disseminate PDEV largely in the NLP and Semantic Web communities. It is distributed in an *Attribution-ShareAlike Creative Commons licence* and is available at `http://pdev.org.uk/PDEVLEMON.html`.

[9] `http://clg.wlv.ac.uk/projects/DVC/`

[10] `http://nlp2rdf.org/nif-1-0`

Acknowledgements

This work was supported by an AHRC grant [Disambiguating Verbs by Collocation project, AH/J005940/1, 2012-2015].

5. References

Beryl T. S. Atkins. 1993. Tools for computer-aided corpus lexicography: the hector project. *Acta Linguistica Hungarica*, 41.

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics.

Paul Buitelaar, Philip Cimiano, John McCrae, Elena Montiel-Ponsada, and Thierry Declerck. 2011. Ontology lexicalisation: The lemon perspective. In Monique Slodzian, Mathieu Valette, Nathalie Aussenac-Gilles, Anne Condamines, Nathalie Hernandez, and Bernard Rothenburger, editors, *Workshop Proceedings of the 9th International Conference on Terminology and Artificial Intelligence*, pages 33–36, Paris, France, November. INALCO.

Philipp Cimiano, Paul Buitelaar, John McCrae, and Michael Sintek. 2011. Lexinfo: A declarative model for the lexicon-ontology interface. *Web Semantics journal*, 9(1):29–51.

Ismaïl El Maarouf, Vít Baisa, Jane Bradbury, and Patrick Hanks. 2014. isambiguating verbs by collocation: Corpus lexicography meets natural language processing. In *Proceedings of LREC*.

Charles J. Fillmore. 1985. Frames and the semantics of understanding. *Quaderni di Semantica*, 6(2):222–254.

Michael A. K. Halliday. 1994. *An introduction to Functional Grammar*. Edward Arnold.

Patrick Hanks and James Pustejovsky. 2005. A pattern dictionary for natural language processing. *Revue Franaise de Linguistique Appliquee*.

Patrick Hanks. 1994. Linguistic norms and pragmatic exploitations, or why lexicographers need prototype theory and vice versa. In F. Kiefer, G. Kiss, and J. Pajzs, editors, *Papers in Computational Lexicography: Complex '94*. Hungarian Academy of Sciences.

Patrick Hanks. 2004. Corpus pattern analysis. In *Euralex Proceedings*, volume 1, pages 87–98.

Patrick Hanks. 2013. *Lexical Analysis: Norms and Exploitations*. MIT Press.

Elisabetta Jezek and Francesca Frontini. 2010. From pattern dictionary to patternbank. In Gilles-Maurice de Schryver, editor, *A Way with Words: Recent Advances in Lexical Theory and Analysis*. Menha Publichers, Kampala.

Adam Kilgarriff and Pavel Rychly. 2010. Semi-automatic dictionary drafting. In Gilles-Maurice de Schryver, editor, *A Way with Words: Recent Advances in Lexical Theory and Analysis*. Menha Publichers, Kampala.

Ken Litkowski. 2012. Corpus pattern analysis of prepositions. Technical report, Damascus, MD: CL Research.

John Mccrae, Guadalupe Aguado-De-Cea, Paul Buitelaar, Philipp Cimiano, Thierry Declerck, Asunción Gómez-Pérez, Jorge Gracia, Laura Hollink, Elena Montiel-Ponsoda, Dennis Spohr, and Tobias Wunner. 2012. Interchanging lexical resources on the semantic web. *Language Resources Evaluation journal*, 46(4):701–719.

Roberto Navigli. 2009. Word sense disambiguation: A survey. *ACM Comput. Surv.*, 41(2):10:1–10:69.

James Pustejovsky. 1995. *The Generative Lexicon*. MIT Press.

Irene Renau and Paz Battaner. 2012. Using cpa to represent spanish pronominal verbs in a learner's dictionary. In *Proceedings of Euralex*, Oslo, Norway.

John Sinclair. 1987. Grammar in the dictionary. In John Sinclair, editor, *Looking up : an account of the COBUILD Project in lexical computing and the development of the Collins COBUILD English language dictionary*, pages 104–115. Collins ELT, London.

John Sinclair. 1991. *Corpus, concordance, collocation*. Oxford University Press.

John Sinclair. 2004. *Trust The Text : Language, Corpus and Discourse*. Routledge.

John Sinclair. 2010. Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins, editors, *In Memory of J.R. Firth*, pages 410–431. Longman, London.

Yorick Wilks. 1975. A preferential, pattern-seeking, semantics for natural language inference. *Artif. Intell.*, 6(1):53–74.