

# Software and Data for Corpus Pattern Analysis

Vít Baisa<sup>1</sup>, Ismaïl El Maarouf<sup>2</sup>, Pavel Rychlý<sup>1</sup>, Adam Rambousek<sup>1</sup>

<sup>1</sup> Natural Language Processing Centre,  
Faculty of Informatics, Masaryk University  
Botanická 68a, 602 00 Brno, Czech Republic  
{xbaisa,pary,rambousek}@fi.muni.cz

<sup>2</sup> Oxford University Press,  
Oxford, United Kingdom  
elmaarouf.ismail@gmail.com

**Abstract.** This report describes the tools and resources developed to support Corpus Pattern Analysis (CPA)—a corpus-based method for building patterns dictionaries. The tools are an annotation of concordance in Sketch Engine, a special CPA editor for editing Pattern Dictionary of English Verbs (PDEV), dedicated servlets based on the Dictionary Editing and Browsing platform and a public interface for browsing the PDEV. The resources are SemEval 2015 Task 15 dataset and LEMON API.

**Keywords:** Corpus Pattern Analysis, Pattern Dictionary of English Verbs, Sketch Engine, linked open data, ontology, LEMON

## 1 Introduction

In this report we present the suite of tools and datasets developed to support the construction of the Pattern Dictionary of English Verbs (PDEV). PDEV is the main output of Corpus Pattern Analysis (CPA), a novel technique in corpus linguistics to map meaning of words onto their patterns of use as observed in real texts. Section 2 gives a brief overview of CPA, Sections 3, 4, 5, 6 present the tools and interfaces used by CPA lexicographers. Sections 7 and 8 introduce a recent work in using the lexicographical resources for NLP. The Bibliography section is a comprehensive compilation of all major publications related to CPA and PDEV<sup>3</sup> not necessarily related directly to this report.

## 2 Corpus Pattern Analysis

Corpus Pattern Analysis (CPA) is a procedure in corpus linguistics which associates word meaning with word use by means of analysis of phraseological

---

<sup>3</sup> And its Spanish and Italian counterparts developed by Irene Renau's team at Pontifical Catholic University of Valparaíso, Chile available online [www.verbario.com](http://www.verbario.com) and by Elisabetta Jezek's team at the University of Pavia, Italy, respectively.

patterns and collocations. In CPA, no attempt is made to identify the meaning of a verb or noun directly, as a word in isolation. Instead, meaning is associated with prototypical sentence contexts. Concordance lines are grouped into semantically motivated syntagmatic patterns. Associating meaning with each pattern is a secondary step, carried out in close coordination with the assignment of concordance lines to patterns. The identification of a syntagmatic pattern is not an automatic procedure: it calls for a great deal of lexicographic art. Among the most difficult of all lexicographic decisions is the selection of an appropriate level of generalization on the basis of which senses are to be distinguished. For example, one might say that the in-transitive verb *abate* has only one sense (to become less in intensity), or one might separate *a storm abates* from *a political protest abates*, on the grounds that the two contexts have different implicatures.<sup>4</sup>

A large apparatus of linguistic categories has been progressively developed to capture corpus patterns for verbs. Patterns can be described according to five types of arguments: Subject, Object, Complement, Adverbial, and Indirect Object. Each can be further specified using determiners, semantic types, contextual roles<sup>5</sup>, and lexical sets. Determiners are used to account for distinctions between “take place” and “take his place”. Semantic types account for distinctions such as “building [[Machines]]” and “building [[Relationship]]”. Contextual roles account for distinctions such as “[[Human = Film Director]] shoot” and “[[Human = Sports Player]] shoot”. Lexical sets account for distinctions such as “reap the whirlwind” and “reap the harvest”.

PDEV is maintained with three main tools: Sketch Engine [1], the CPA editor and the DEB server. The corpus used is the BNC [2], a large reference corpus containing various text types in British English (100 million words). For the purpose of the CPA analysis, the corpus has been filtered and only its written part was used. The result is usually referred as BNC50 since it contains roughly 50 million words. Lexicographers extract typical phraseological patterns from corpora by clustering corpus tokens (labelling them) according to the similarity of their context.

### 3 Annotation in Sketch Engine

Sketch Engine supports the annotation of tokens in any corpus using unique identifiers that refer to labels manually defined by lexicographers. The starting point is the creation of a concordance, based on a lemma and part-of-speech tag (lempos). In the web interface [3], the annotation is facilitated by green boxes next to each KWIC (Key Word In Context). The set of labels consists of numbers (for pattern numbers) although words can be used instead and each label can be further decomposed into sub-labels indicating a variation of use

---

<sup>4</sup> The implicature is a term from pragmatics referring to what is suggested in an utterance.

<sup>5</sup> Semantic types and roles are enclosed in double square brackets, roles are separated by “=”.

(‘.a’ for anomalous argument, ‘.f’ for figurative use of a pattern, and ‘.s’ for syntactically anomalous). Label ‘x’ is used for tagging errors (e.g. an adjective use for a verb token) and ‘u’ for an unclassifiable word occurrence. Moreover, a label can be assigned to a whole page, or to a set of selected lines. The set of labels can be modified at any time.

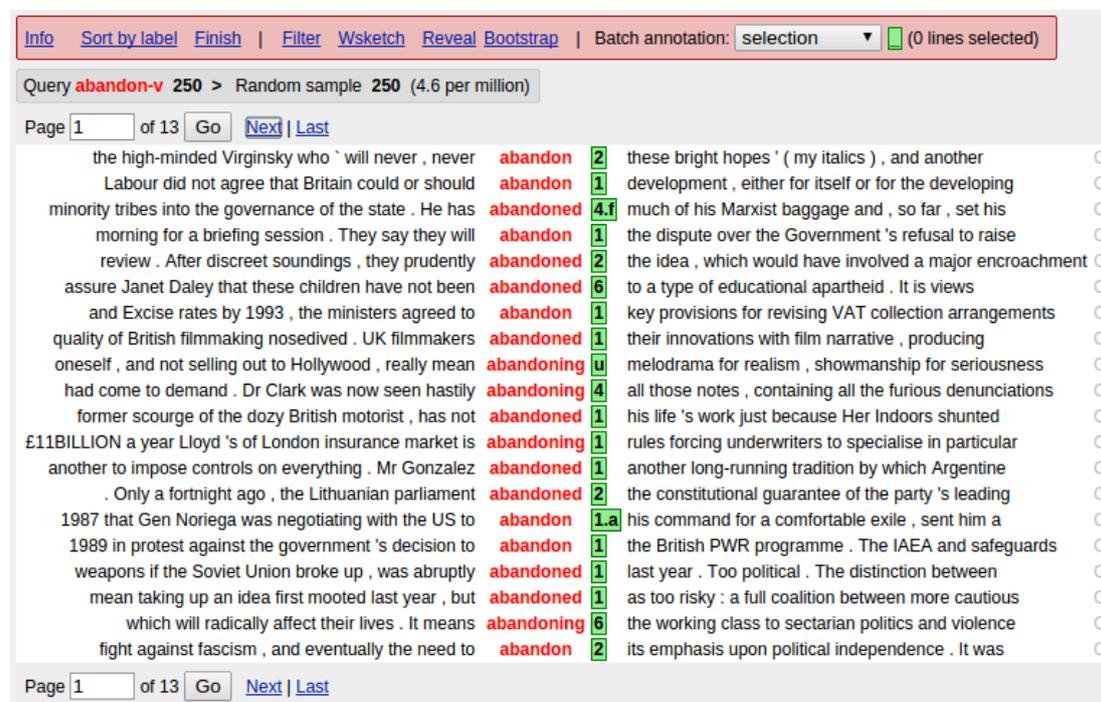


Fig. 1. The annotation interface in Sketch Engine.

Labels are stored as IDs internally which allows for renaming labels in batches. The mapping files are stored one per concordance. It is also possible to 1) undo an action, 2) sort the concordance according to the label values, 3) combine with standard sorting (by left or right context), 4) visualize the annotation statistics (Figure 2), 5) filter concordance by labels or 6) annotate via word sketches.

The labelling actions are the part of Bonito API<sup>6</sup>: each assignment of labels to line(s) is sent via AJAX to the server and saved. The API has methods for assigning labels, renumbering labels (changing the mapping globally for an annotation), locking the current annotation, obtaining annotation statistics etc.

A training mode is also available to enable trainees to annotate a concordance sample without altering the existing annotations. In this mode each annotation by a trainee is stored separately and can be viewed by a trainer. The performance of each trainee with respect to the master concordance can also be

<sup>6</sup> [www.sketchengine.co.uk/json-api-documentation](http://www.sketchengine.co.uk/json-api-documentation)

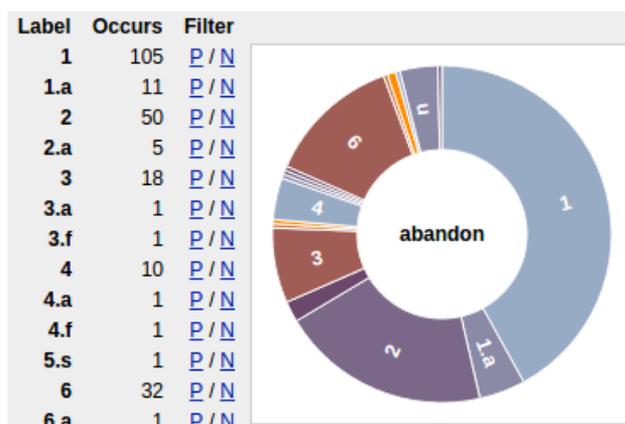


Fig. 2. Visualization of annotation statistics for verb “abandon”.

computed. Finally a trainee’s concordance can be validated by the trainer and merged with the main concordance.

Other features have been developed such as an automatic clustering of concordance lines, a bootstrapping mechanism to annotate the rest of a partially annotated concordance using collocation-based features identified in the context of the lines already annotated. The evaluation of the bootstrap was given in [4].

## 4 CPA editor

The CPA editor is implemented in JavaScript (using jQuery and DataTables) and PHP. It mainly consists of an entry manager (Figure 3) and a pattern editor (Figure 4). The entry manager retrieves information about entries and stores them in a DataTable table, thus enabling easy sorting of columns and filtering using the search input box or verb statuses (complete, ready, work in progress). Each line of the table can be selected/unselected to be printed in CSV. When a user clicks on an entry, the entry manager opens an instance of a pattern editor in a new tab, which is the workspace for the entry.

The pattern editor is organized in three main parts. 1) The top bar contains information for the entry such as the sample size, the status or comments. 2) The pattern list displays patterns and their frequency in the annotated concordance, the pattern string, the implicature and also indications about whether the pattern is an idiom or whether it is used in a particular register. 3) Clicking on any of the pattern line opens a pattern box (Figure 5). This box lists arguments line by line and each line can be expanded to account for alternating components (such as the fact that two semantic types are equally valid in a given argument). Various information can be added in the form of check boxes, menu lists, etc. The final line is the implicature, which can be automatically generated based on the pattern string which adds up all the elements in the pattern.

**Pattern Dictionary of English Verbs** Expand SKE Print patterns Unselect Ontology Listing Report a bug Print query

Filter:  all Show/Hide columns: Verb Status Patterns Sample BNC50 BNC OEC FN Links Created by Created Last editor Modified Print

Verb	Status	Patterns	Sample	BNC50	BNC	OEC	FN Links	Created by	Created	Last editor	Modified	Print
break	ready	83	2000	8297	18603	186711	13	patrick	2009/06-01	saramoze	2015-08-25	<input type="checkbox"/>
blow	complete	62	1516	1516	4796	55320	6	patrick	2009/06-01	patrick	2015-01-26	<input type="checkbox"/>
throw	complete	61	1000	3710	10919	143403	7	patrick	2007/06-01	ymaarouf	2015-03-18	<input type="checkbox"/>
lose	ready	57	1000	11868	26605	301942	2	jane	2014/12-08	saramoze	2015-01-25	<input type="checkbox"/>
take	WIP	56	1000	75872	173412	1733310	5	patrick	2009/11-26	saramoze	2015-05-13	<input type="checkbox"/>
open	ready	56	1000	8695	22394	268691	0	cpa04	2006/11-01	jane	2014-12-15	<input type="checkbox"/>
go	G	50	54872	54872	226268	2127417	7	patrick	2008/06-20	jane	2014-12-05	<input type="checkbox"/>
live	ready	43	1000	15402	31991	316892	1	patrick	2009/06-01	jane	2014-12-05	<input type="checkbox"/>
set	G	38	20542	20542	38838	315361	0	patrick	2008/05-05	jane	2014-11-18	<input type="checkbox"/>
hit	WIP	38	1000	3706	10344	173794	0	patrick	2006/09-30	jane	2014-12-15	<input type="checkbox"/>
hang	ready	38	500	2242	8659	96907	11	patrick	2009/06-01	patrick	2014-12-20	<input type="checkbox"/>
beat	ready	37	1000	2224	7859	101890	2	patrick	2007/07-19	imaarouf	2015-09-18	<input type="checkbox"/>
call	complete	36	1000	24439	51912	591606	12	patrick	2006/09-30	patrick	2014-12-05	<input type="checkbox"/>
dig	ready	30	845	845	2623	27870	0	jezek	2007/05-26	jane	2014-12-05	<input type="checkbox"/>

Showing 1 to 5,601 of 5,601 entries

Fig. 3. CPA editor – the entry manager.

**owe** Add pattern Stretch Shrink more Concordance Ontology Renumber Save Save&Close Close

Sample size  (out of 2026) Semantic class  Status  Difficulty  Compilation time

#	%	Pattern & primary implicature
1.	25.00%	[[Human 1   Institution 1]] owe [[Human 2   Institution 2]] [[Money]] (for [[Physical_Object]]   for [[Asset]]) [[Human 1   Institution 1]] is under obligation to repay [[Money]] borrowed from [[Human 2   Institution 2]]
2.	19.67%	[[Human 1]] owe [[Human 2   Institution]] [[Obligation]] [[Human 1]] is morally and/or legally bound to honour [[Obligation]] to [[Human 2]]
3.	17.00%	[[Entity]] owe [[REFLDET Privilege   REFLDET Property   REFLDET (Eventuality = Desirable)]] (to [[Anything]]) [[Entity]] is able to gain [[Privilege   Property]], or have [[Eventuality = Desirable]] happen to them because of [[Anything]]
4.	13.67%	[[Entity]] owe {much   little   ...} (to [[Anything]]) [[Human   Institution   Concept]] is able to develop intellectually, culturally, economically or otherwise because of [[Anything]]
5.	15.00%	[[Eventuality 1]] owe {much   something   ...} (to [[Eventuality 2]]) [[Eventuality 1]] is, to a certain extent, caused or affected by [[Eventuality 2]] [[Anything]] either contributed to, or partially caused [[Eventuality]] to take place
6.	2.33%	[[Human 1]] owe {it} (to [[Self]]   to [[Human 2]]) (to/INF [V]) [[Human 1]] feels morally obligated to do something for [[Self   Human 2]]
7.	2.33%	[[Human 1]] owe [[Human 2]] {a debt (of gratitude)} idiom [[Human 1]] feels thankful to [[Human 2]]
8.	1.00%	[[Human 1]] owe [[Human 2]] {apology} [[Human 1]] needs to apologize to [[Human 2]]

Fig. 4. CPA editor – the pattern manager.

**Pattern gnaw 4** In corpus:  Concordance  Insert Merge into:  Copy Delete Save Save&Close Close

Subject   Lexset

Verb  no object  no adverbial

Adverbial      Lexset

Opt

Primary implicature   
[Generate](#)

idiom  pv

Show: Sub. conjunction Indirect object Object Complement Clausals Clausals objects Secondary implicature Domain & Register  
Framenet Comment Sem. Class

Fig. 5. Pattern box in the CPA editor.

The pattern editor contains many features, here we mention only a few: 1) adding a new pattern, 2) renumbering the list of patterns (by drag & drop), 3) inserting a new pattern at a specific place, 4) merging two patterns, 5) accessing a filtered concordance of all the lines tagged with a specific pattern, etc. The editor is synchronized with Sketch Engine annotations periodically, such that every time a pattern is added or removed, the information is updated in both systems. The pattern editor comes with an ontology editor which enables the lexicographer to create, delete or update semantic type nodes in the CPA ontology. Several methods have been developed in order to access all the verbs which use a particular semantic type, and all the nouns matching a semantic type in the corpus (Figure 6).

The screenshot shows the 'Ontology' editor interface. On the left, a tree view displays the hierarchy of semantic types: Anything (i.e. anything at all), Entity, Abstract\_Entity, Concept (Must be a word meaning 'concept'; otherwise use Anything = Concept), Proposition, Narrative, Rule, Permission, Dispute, Information, Information\_Source, Document (Information\_Source, Artifact), Agreement (Speech\_Act, Document), and Language. On the right, there are search boxes for 'Semantic Types' and 'Nouns'. Below the search boxes, a table displays the results for the semantic type 'ATTITUDE'. The table has columns for Verb, Pattern number, Freq, and Nouns (with sub-columns S, O, Ad, Σ).

Verb	Pattern number	Freq	Nouns	S	O	Ad	Σ
arouse	1,2	763	glory	0	0	15	15
resent	1	457	intention	0	10	0	10
applaud	3	158	determination	0	9	0	9
awaken	3,4,5	78	glow	0	0	7	7
brush	1,5	76	willingness	0	7	0	7
repeat	1	72	stance	0	6	0	6
trigger	2	55	acceptance	0	4	0	4
greet	2	50	attitude	0	3	0	3
admit	6	39	readiness	0	3	0	3

Fig. 6. Ontology editor with populated nouns and patterns.

The JavaScript code of the editor was also used for Pattern Dictionary of English Prepositions by Ken Litkowski [5].

## 5 DEB platform

The CPA editor interacts with the database server via API (Ruby, WebBrick) which manages the CPA databases (one per language, currently English, Italian and Spanish). The server is based on the DEB<sup>7</sup> dictionary management server [6] and is based on private and public methods. Several servlet methods have been designed in order to ensure maximum efficiency as well as to limit server overload. The general mechanism is that a JavaScript client sends AJAX queries to the DEB server which queries the database and sends back responses in JSON. There are two main servlets, one for the CPA editor, one for the public version (Section 6). The server also maintains users and their privileges (e.g. only expert users can label a verb as complete in the editor, or modify them thereafter).

<sup>7</sup> deb.fi.muni.cz

## 6 Public access to PDEV

To provide access to the up-to-date data of PDEV and the ontology we have developed a user-friendly online tool<sup>8</sup> which is connected to the main PDEV database. It is structured similarly to the editor but the style has been entirely re-shaped and only a limited number of methods are available (it does not include pattern boxes for instance). The website uses colour codes for different pieces of information (semantic types, grammatical categories, lexical items). The public website interacts with Sketch Engine to access labelled examples for each pattern. A specific feature enables to show the best sentence example (GDEX, [7]) for each pattern in the pattern list (Figure 7). Ontology in the form of a semantic type list and a hierarchical structure is also available.

The screenshot displays the 'Pattern Dictionary of English Verbs' interface. At the top, there are logos for the University of Wolverhampton and NLP. The main navigation bar includes links for 'About CPA', 'Browse Verbs', 'The Sketch Engine', 'Publications', 'CPA Ontology', 'Semantic Types', 'Download', and 'report a problem'. Below the navigation bar, there are search filters for 'Browse: complete verbs (1286)', 'work-in-progress verbs (443)', 'not yet started verbs (3667)', and 'all verbs (5396)'. A search box with a 'Find a verb' button is present. The main content area shows 'PDEV: argue' with an 'Access full data' button. To the right, there are controls for 'Displayed here are All patterns' and 'sample size: 250 patterns: 7'. The list of patterns includes:

Pattern	Implicature	Example	Percentage
1 Pattern: Human or Institution or Document argues QUOTE or THAT-CLAUSE	Human or Institution or Document states reasons for believing [CLAUSE]	The country's nuclear lobby has argued that alternative energy sources are either not available or too expensive	88.4%
2 Pattern: Human argues Proposition QUOTE or THAT-CLAUSE	Human states reasons for believing Proposition	a landscape architect was arguing the case for the railroad companies to plant station gardens to advertise both the train service and the town it served.	2.0%
3 Pattern: Human or Institution or Document argues for or in favour of Action	Human or Institution or Document states reasons in favour of doing Action	Various authors have argued for seasonal camps and settlements based on the animal resources available	3.2%
4 Pattern: Human or Institution or Document argues against Action	Human or Institution or Document states reasons in favour of not doing Action	many conservationists have argued against the commercial production of timber	2.4%

Fig. 7. Public interface to PDEV data.

## 7 SemEval 2015 Task 15 dataset

In order to support NLP research in semantic parsing which would help to evaluate the impact of the CPA resources in semantic tasks, a high quality dataset derived from PDEV was produced and used in Task 15 at Semeval competition in 2015<sup>9</sup>. The goal of this task was to evaluate to which extent NLP systems could contribute to the creation of a lexicographical entry. To maximize participants' interest as well as to simplify this complex task, it was broken down into 3 inter-connected subtasks: 1) CPA parsing: all sentences in the dataset to be syntactically and semantically parsed. 2) CPA clustering: all sentences in the dataset to be grouped according to their similarities. 3) CPA pattern editing: all verb patterns found in the dataset to be described in terms of their syntactic and semantic properties.

<sup>8</sup> [www.pdev.org.uk](http://www.pdev.org.uk)

<sup>9</sup> [alt.qcri.org/semEval2015](http://alt.qcri.org/semEval2015)

Two datasets were created: Microcheck, which included all three tasks, and was intended to be used in analysing the correlation of the tasks; Wingspread, a larger dataset yet only including task 2 and 3, as task 1 required manual annotation of semantic and syntactic properties of arguments in the context of the verb. It is worth noting that this was the first attempt at annotating CPA pattern arguments in context. For more details, refer to SemEval paper [8]. The resource is available from the SemEval website<sup>10</sup>.

While the task did not gather as much interest as desired, it fostered the development of a baseline system (mostly unbeaten by competing systems), which connected all three tasks together. This system was integrated into the CPA interface under the username “auto-cpa” for scrutiny and validation by CPA lexicographers.

## 8 LEMON API

It is important to release the data produced by CPA lexicographers for use by NLP developers, in the same fashion as WordNet, FrameNet and other resources.

The problem for CPA is that the data was scattered in different files, in different formats, spanning several database tables, each holding heterogeneous types of content (corpus examples, links, ontology, ...).

The solution was to encode the resource using RDF as linked open data in a DEB server method having access to Sketch Engine annotation data. The name of this project is “PDEV-LEMON” and the first version was released in 2014 [9]. The server script connects elements of patterns stored in the database to the CPA ontology, and calls Sketch Engine API methods to retrieve annotated examples. The dictionary is encoded using the LEMON model [10] which provides the general structure and features to enable an easy instantiation of a lexicon using an ontology framework such as OWL. PDEV-LEMON includes 7 ontologies which describe the CPA semantic ontology as well as ontologies describing specific concepts and relations used in PDEV-LEMON, taxonomies of domains, and registers and so on. The first release of PDEV-LEMON included 17,634 triples, 3,702 patterns and 10,799 arguments.

A second release is planned for the end of 2015 and it will include a full linked data encoding of examples (which, in the first release linked back to the public access website) using NIF.<sup>11</sup> The conversion script was made in such a way as to handle other languages (Spanish and Italian so far), so it is expected that future releases will include dictionaries in languages other than English.

## 9 Conclusion

This report describes the software tools to support the development of pattern dictionaries applying the CPA method. Latest developments of the CPA

<sup>10</sup> [alt.qcri.org/semeval2015/task15](http://alt.qcri.org/semeval2015/task15)

<sup>11</sup> NLP Interchange Format: [persistence.uni-leipzig.org/nlp2rdf](http://persistence.uni-leipzig.org/nlp2rdf)

infrastructure have made it easier for lexicographers to quickly draft high quality dictionary entries, while being robust for a large number of simultaneous users. Several projects, such as PDEV-LEMON, and the development of the Microcheck and Wingspread datasets, have been launched to disseminate the resource and facilitate its use in NLP.

In parallel, this infrastructure was (and is being) also used to develop verb pattern dictionaries for Spanish<sup>12</sup> and Italian. In the future, if the number of languages increases (particularly with Czech, German and French), it should facilitate the painstaking lexicographical work needed to develop pattern dictionaries.

Terminology and language learning resources can also benefit from this suite of tools. Among future improvements of the CPA editor, we may mention providing support to connect patterns monolingually and across languages (English, Spanish, Italian). These links could substantially contribute to areas such as knowledge-based Machine Translation.

The CPA editor and the annotation in Sketch Engine are used on daily basis by lexicographers.

**Acknowledgement** We would like to thank lexicographers, annotators, trainees, testers, who have used these tools and provided feedback and to Patrick Hanks, the main user of these tools, for his unrelenting input and patience.

This work has been partly supported by an AHRC grant [DVC, AH/J005940/1, 2012–2015], and by the Czech Science Foundation under the project GA15-13277S, and by the project MUNI/A/1165/2014 *Čeština v jednotě synchronie a diachronie* – 2015.

## References

1. Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V.: The sketch engine: ten years on. *Lexicography* 1(1) (2014) 7–36
2. Leech, G.: 100 million words of english: the british national corpus (bnc). *Language Research* 28(1) (1992) 1–13
3. Rychlý, P.: Manatee/bonito-a modular corpus manager. In: 1st Workshop on Recent Advances in Slavonic Natural Language Processing, Faculty of Informatics (2007) 65–70
4. El Maarouf, I., Bradbury, J., Baisa, V., Hanks, P.: Disambiguating verbs by collocation: Corpus lexicography meets natural language processing. In: Proceedings of LREC, Reykjavik, Iceland (2014) 1001–1006
5. Litkowski, K.: Pattern dictionary of english prepositions. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, Maryland, Association for Computational Linguistics (June 2014) 1274–1283

<sup>12</sup> [verbario.com](http://verbario.com)

6. Horák, A., Vossen, P., Rambousek, A.: A distributed database system for developing ontological and lexical resources in harmony. In: *Computational Linguistics and Intelligent Text Processing*. Springer (2008) 1–15
7. Kilgarriff, A., Husák, M., McAdam, K., Rundell, M., Rychlý, P.: Gdex: Automatically finding good dictionary examples in a corpus. In: *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. (2008) 425–432
8. Baisa, V., Bradbury, J., Cinková, S., El Maarouf, I., Hanks, P., Kilgarriff, A., Popescu, O.: Semeval-2015 task 15: A cpa dictionary-entry-building task. In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Denver, Co, USA (2015)
9. El Maarouf, I., Bradbury, J., Hanks, P.: Pdev-lemon: a linked data implementation of the pattern dictionary of english verbs based on the lemon model. In: *Proceedings of the 3rd Workshop on Linked Data in Linguistics (LDL): Multilingual Knowledge Resources and Natural Language Processing at the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland (2014)
10. McCrae, J., Aguado-de Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., et al.: Interchanging lexical resources on the semantic web. *Language Resources and Evaluation* **46**(4) (2012) 701–719
11. Hanks, P.: How people use words to make meanings: Semantic types meet valencies. In Boulton, A., Thomas, J., eds.: *Input, Process and Product: Developments in Teaching and Language Corpora*. Masaryk University Press, Brno (2012) 54–69
12. Kilgarriff, A., Palmer, M.: Introduction to the special issue on senseval. *Computers and the Humanities* **34**:1–2 (2000)
13. Hanks, P., Pustejovsky, J.: A pattern dictionary for natural language processing. *Revue Française de linguistique appliquée* **10**:2 (2005)
14. Hanks, P.: *Lexical Analysis: Norms and Exploitations*. MIT Press, Cambridge, MA (2013)
15. Jezek, E., Magnini, B., Feltracco, A., Bianchini, A., Popescu, O.: T-pas; a resource of typed predicate argument structures for linguistic analysis and semantic processing. In: *Proceedings of LREC, Iceland (2014)*
16. Renau, I., Battaner, P.: Using cpa to represent spanish pronominal verbs in a learner's dictionary. In: *Proceedings of the XV EURALEX, Norway (2012)*
17. Hanks, P.: Corpus pattern analysis. In: *Proceedings of the XI EURALEX, Lorient, France (2004)*
18. Alonso Campo, A., Renau, I.: Corpus pattern analysis in determining specialised uses of verbal lexical units. *Terminalia* **7** (2013) 26–33
19. El Maarouf, I., Baisa, V.: Automatic classification of semantic patterns from the pattern dictionary of english verbs. In: *Proceedings of JSSP2013, Trento, Italy (2013)* 95–99
20. Bradbury, J., El Maarouf, I.: An empirical classification of verbs based on semantic types: the case of the 'poison' verbs. In: *Proceedings of JSSP2013, Trento, Italy (2013)* 70–74
21. Cinková, S., Holub, M., Kríž, V.: Optimizing semantic granularity for nlp - report on a lexicographic experiment. In: *Proceedings of the 15th EURALEX International Congress, Oslo, Norway (2012)* 523–531
22. Mills, C., Levow, G.A.: Cmills: Adapting srl features to dependency parsing. In: *Proceedings of SemEval 2015, Denver, USA (2015)*
23. Pedersen, T.: Duluth: Word sense discrimination in the service of lexicography. In: *Proceedings of SemEval 2015, Denver, USA (2015)*

24. Feng, Y., Deng, Q., Yu, D.: Blcunlp: Corpus pattern analysis for verbs based on dependency chain. In: Proceedings of SemEval 2015, Denver, USA (2015)
25. Cinková, S., Holub, M., Kríž, V.: Managing uncertainty in semantic tagging. In: Proceedings of 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France (2012) 840–850
26. Cinková, S., Holub, M., Rambousek, A., Smejkalová, L.: A database of semantic clusters of verb usages. In: Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey (2012) 3176–3183
27. Holub, M., Kríž, V., Cinková, S., Bick, E.: Tailored feature extraction for lexical disambiguation of english verbs based on corpus pattern analysis. In: Proceedings of the 24th International Conference on Computational Linguistics (Coling 2012), Mumbai, India (2012) 1195–1209
28. Kawahara, D., Peterson, D.W., Popescu, O., Palmer, M.: Inducing example-based semantic frames from a massive amount of verb uses. In: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics. (2014) 58–67
29. Popescu, O.: Building a resource of patterns using semantic types. In: Proceedings of LREC, Boston, USA (2012)
30. Popescu, O.: Learning corpus patterns using finite state automata. In: Proceedings of the 10th International Conference on Computational Semantics, Potsdam, Germany (2013) 191–203
31. Popescu, O., Palmer, M., Hanks, P.: Mapping cpa onto ontonotes senses. In: Proceedings of LREC, Reykjavik, Iceland (2014) 882–889
32. Pustejovsky, J., Hanks, P., Rumshisky, A.: Automated induction of sense in context. In: Proceedings of COLING, Geneva, Switzerland (2004)
33. Rumshisky, A., Hanks, P., Havasi, C., Pustejovsky, J.: Constructing a corpus-based ontology using model bias. In: Proceedings of FLAIRS, Melbourne, FL (2006) 327–332
34. Sinclair, J.: *Corpus, concordance, collocation*. Oxford University Press (1991)
35. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The berkeley framenet project. In: Proceedings of the 17th international conference on Computational linguistics-Volume 1, Association for Computational Linguistics (1998) 86–90
36. Pustejovsky, J.: *The Generative Lexicon*. MIT Press (1995)
37. Fillmore, C.J.: Frames and the semantics of understanding. *Quaderni di Semantica* 6(2) (1985) 222–254
38. Wilks, Y.: A preferential, pattern-seeking, semantics for natural language inference. *Artif. Intell.* 6(1) (1975) 53–74
39. Kilgarriff, A., Rychlý, P.: Semi-automatic dictionary drafting. In de Schryver, G.M., ed.: *Oxford Handbook of Innovation*. Menha Publishers, Kampala (2010)
40. Litkowski, K.: *Corpus pattern analysis of prepositions*. Technical report, CL Research (02 2012)
41. El Maarouf, I., Alferov, E., Cooper, D., Fang, Z., Mousselly Sergieh, H., Wang, H.: The guanxi network: a new multilingual llod for language learning applications. In: Proceedings of NLP&LOD workshop, RANLP. (2015)
42. El Maarouf, I., Marsic, G., Orasan, C.: Barbecued opakapaka: Using semantic preferences for ontology population. In: Proceedings of RANLP. (2015)
43. Jezek, E., Hanks, P.: What lexical sets tell us about conceptual categories. *Lexis* 4(7) (2010) 22
44. Pustejovsky, J., Jezek, E.: Semantic coercion in language: Beyond distributional analysis. *Italian Journal of Linguistics* 20(1) (2008) 175–208

45. Popescu, O., Vo, N.P.A., Feltracco, A., Jezek, E., Magnini, B.: Toward disambiguating typed predicate-argument structures for italian. Proceedings of CLIC-IT14 (2014)
46. Feltracco, A., Jezek, E., Magnini, B.: Opposition relations among verb frames. In: Proceedings of the 3rd Workshop on EVENTS at the NAACL-HLT. (2015) 16–24