PHRASALEX III

PHRASEOLOGICAL APPROACHES TO LEXICOGRAPHY

PHRASEBASE TALK

Introducing PhraseBase: A linguistic information System for language learners, translators and for NLP

Laura Giacomini
Laura Rebosio

universität innsbruck

Overview

1. Structure and scope of the PhraseBase project
2. People
3. Theoretical foundations and development
4. Methodology and data

1. Structure and scope of the PhraseBase project

PhraseBase is a **Linguistic Information System** consisting of three main components,

- a dictionary,
- an ontology/thesaurus and
- a grammar,

primarily for second language acquisition and natural language processing (NLP)

PhraseBase → *phraseological database*

The theoretical framework behind PhraseBase is phraseological & cognitivist → Sinclair‘s theory, Hanks‘s formalisation, DiMuccio-Failla‘s further development

PhraseBase includes a PAD (Phrase-based Active Dictionary) → currently: multi-monolingual dictionaries for IT, DE, EN

Contrastive perspective: search for partial or total equivalence of frames (typical situations) across languages and cultures

Ideal user: advanced learner, translator

PhraseBase

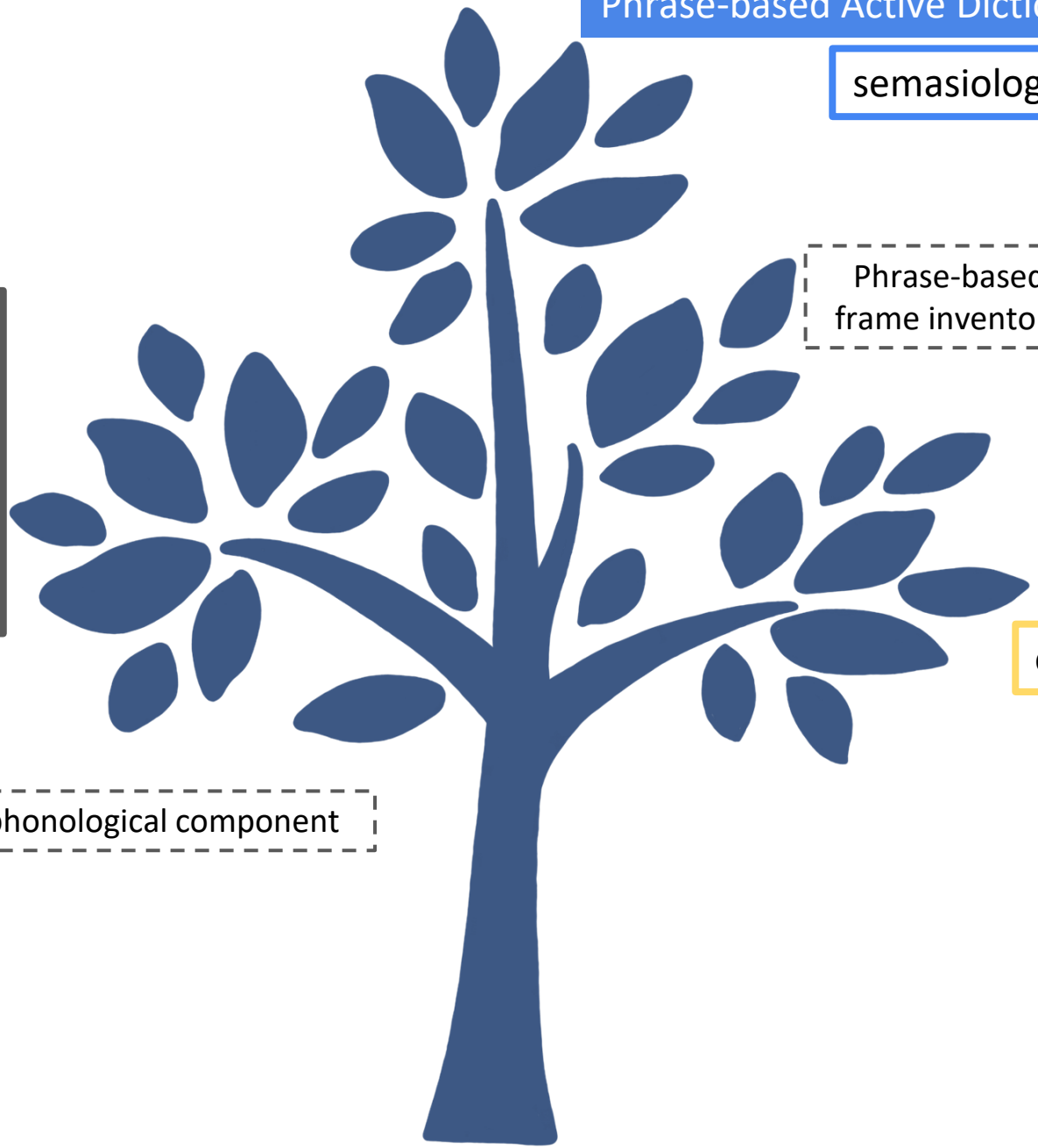
PAD
Phrase-based Active Dictionary

semasiological component

Phrase-based
frame inventory

inventory of
constructions
(constructicon)

grammatical
component



PHRASEnet
Phrase-based Wordnet

onomasiological component

phonological component

2. People

LAURA GIACOMINI (Innsbruck, previously Heidelberg/Hildesheim)	→ PI, project initiator, methodological framework, data modelling, data analysis
PAOLO DI MUCCIO-FAILLA (Hildesheim)	→ project initiator, theoretical and methodological framework, data modelling/programming, data analysis
ADRIANA ORLANDI (Modena and Reggio-Emilia)	→ organiser of PhrasaLex I, first experiments on FR
EVA LANZI (Heidelberg)	→ data analysis
SARAH PIEPKORN (Hildesheim)	→ data analysis, project on aspectuality of verbs
FRITZ KLICHE (Hildesheim)	→ NLP approaches to data analysis
LAURA REBOSIO (Innsbruck)	→ data analysis, project on ostensive, e.g. frame-based definitions
LINDA PROSSLINER (Innsbruck)	→ data analysis, project on idiomatic expressions for children
GIULIANO GIAMBERTONE (Innsbruck)	→ DB/web programming

3. Theoretical foundations and development

Session 1 on Thursday

universität innsbruck

The lexicon of a language is phraseological in nature.

→ Semantic ambiguity can be reduced if one takes in consideration the context in which words are used.

→ Chunks of linguistic expressions – and not single words – are identified as lexical units.

→ Meaning distinctions can be (easily) ascertained because they correspond to *word usage patterns*.

Sinclair (cf. 2004: 133): *not isolated words, but **words in their contextual patterns of normal usage** are the most common lexical units of language*.

Sinclair (1991: 65): *It seems that there is a strong tendency **for sense and syntax to be associated***.

Hanks (cf. 2013: 192): *Normal collocations are statistically significant in a corpus analysis. Asking for the meaning of a word turns out in asking for the **meaning of a pattern**. Words in isolation have only potential meanings.*

Hanks (cf. 2013: 5): *In a better dictionary, it should be listed what is linguistically (semantically) normal and not, what is ever semantically possible. A distinction should be made between **normal meaning variations** and **exploitations**.*

DiMuccio-Failla & Giacomini (2017a,b)

About: ***word usage patterns***

universität innsbruck

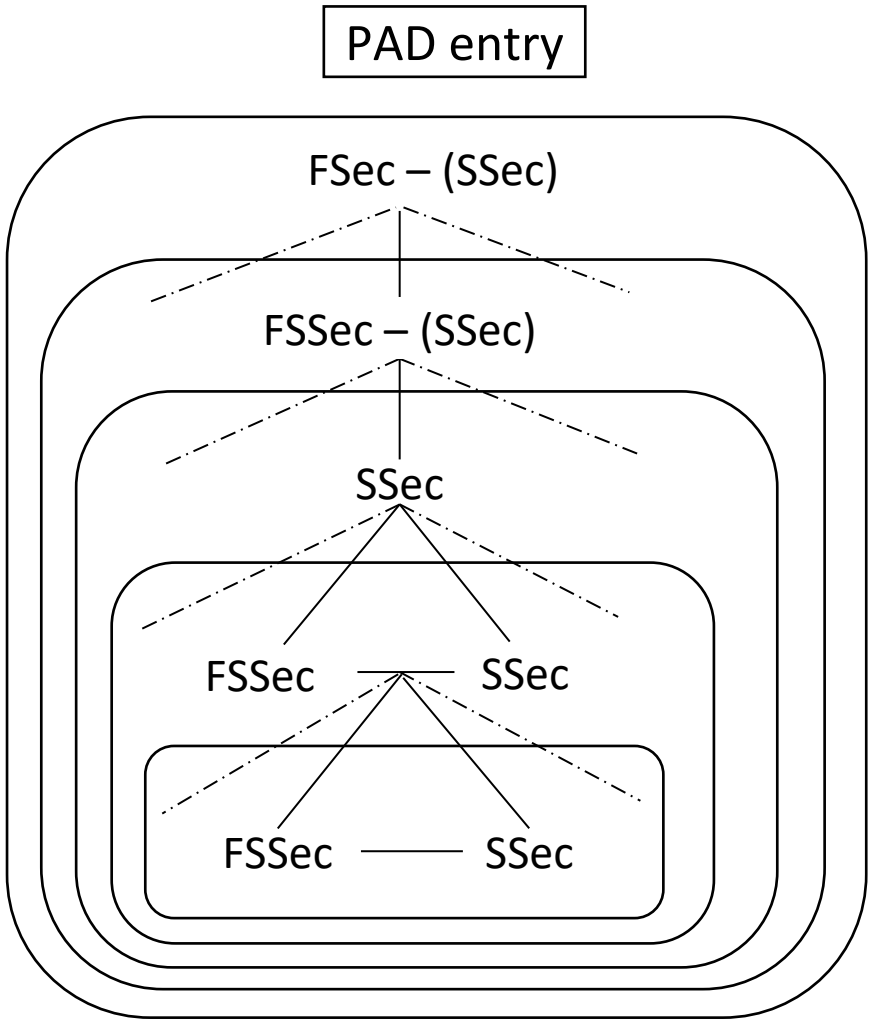
Examples: (1) so. puts sth. in a particular place or position ← *I put my phone in your bag.*
(2) so. puts so. somewhere ← *Dad puts the children to bed.*
(3) so. puts sth. on so. ← *The boss will put extra pressure on you.*

- A ***normal*** *word usage pattern* generally has only one meaning.
- *Normal* means ***typical***; typical, recurring patterns are the most frequent ones in a corpus; a *normal* meaning is the common, conventional meaning associated traditionally with that pattern within a specific linguistic community.
- A normal word usage pattern is determined by **four features**: its collocation, its colligation, its semantic preference and its semantic prosody.
- **Intuition** and **introspection** of the lexicographer are crucial in analyzing the data and evaluating evidence.

The PAD microstructure (1)

DiMuccio-Failla & Giacomini (2022)

FSec = Formal Section
FSSec = Formal-Semantic Section
SSec = Semantic Section



The PAD microstructure (2)

agree \ə ˈgriː\

VERB, REGULAR

to AGREE ‹WITH sb. / s. opinion› ‹ON/ABOUT s.e.› [opinion]

/ABSOLUTELY/TOTALLY/STRONGLY/CERTAINLY / NOT NECESSARILY / NOT QUITE

▶ to think that sb. is right ‹on/about s.e.›

~

▶ to share s. opinion ‹on/about *›

to AGREE (with each other) ON s. decision ‹THAT…›/‹TO do sth.› [expr. of decision]

|TOGETHER | BETWEEN ONESELVES | MUTUALLY [at a particular moment]

▶ ⌜to decide to do sth.⌝ together

~

▶ to choose sth. together

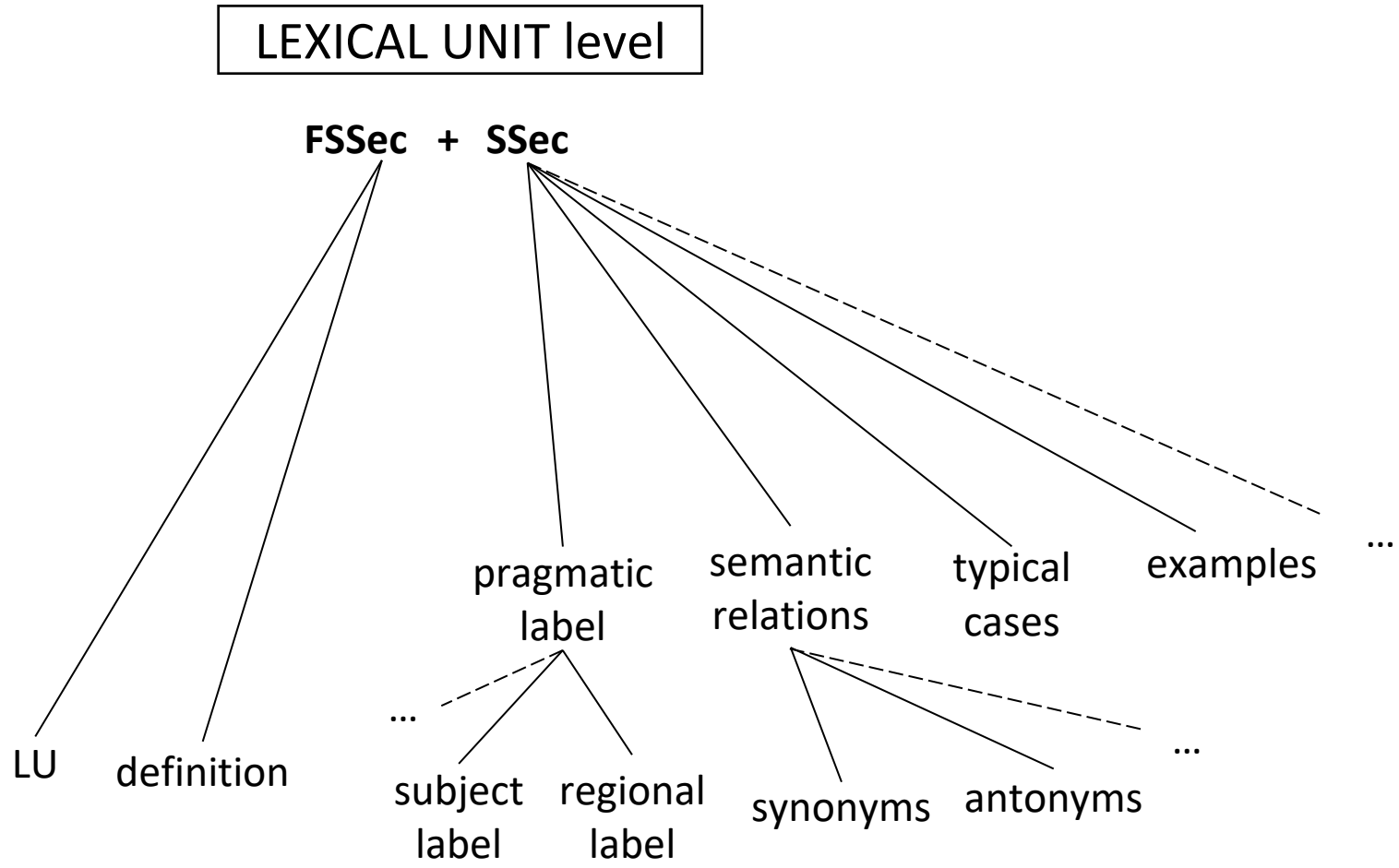
1 to AGREE ‹WITH a given person› …

… ■ **‹ON a given SUBJECT/TOPIC/ISSUE› OR ‹ABOUT a certain entity›** ○ **to think that a given person's opinion/assessment ‹on ·› OR ‹about ·› is right**

• EXAMPLES: ① … • SYNONYMS: ① [FML.] to CONCUR ‹WITH ·› ‹ON/ABOUT ·› ② to SHARE a given person's OPINION ‹ON/ABOUT ·› ③ [FML.] to BE IN AGREEMENT ‹WITH ·› ‹ON/ABOUT ·› ④ to THINK THE SAME ‹AS a given person› ‹ON/ABOUT ·› ⑤ [FML.] to BE OF THE SAME OPINION=MIND ‹AS a given person› ‹ON/ABOUT ·› ⑥ to HAVE THE SAME OPINION ‹AS a given person› ‹ON/ABOUT ·› • ANTONYMS: ① to DISAGREE ‹WITH ·› ‹ON/ABOUT ·› ② [FML.] to NOT SEE EYE TO EYE ‹WITH ·› ‹ON/ABOUT

The PAD microstructure (3)

DiMuccio-Failla & Giacomini (2022)



Cognitivist account on polysemy (cf. DiMuccio-Failla: forthcoming)

- Brugman & Lakoff (1988: 478): in a speaker's mind, the related senses of a word are organised in a radial set around one or more **prototypical concepts**. Each individual sense is a conceptual category organised around prototypical members.

  Example: The central sense of *over* combines elements of both *above* and *across.*

  → The links between the senses are instances of metonyms, metaphors, image-schema transformations, shifts within a semantic frame, ect. The boundaries of a single sense need not to be clear-cut. The lexical network is a **network** of minimally differing senses (Norvig & Lakoff 1987: 195).

- Johnson (1987): **embodiment** of mental concepts: *Image-schemata are structures for organizing our experience and comprehension* (cf. p. 29).

Presentation of word meaning through ostensive aids (PhD project)

universität innsbruck

GOALS: 1) issuing guidelines for a systematic identification of polysemous senses and ordering them in the entry according to semantic-cognitive principles, with 'core/prototypical meaning' first.

→ syntactic constructions and cognitive representations of meaning are often at odds

→ what is the prototypical meaning?

→ can the user easily find the linguistic expression for the concept he/she has in mind? (*active*)

2) presenting word meaning through ostensive aids, e.g. phrase-based pictorial frames

→ what kind of visual aids are suitable for which words?

→ pictures have, like prototypical concepts, no boundaries

→ implementing AI?

4. Methodology and data

Session 2 on Thursday

universität innsbruck

- Gathering collocations from corpora, general dictionaries, collocation dictionaries, …
- Grouping collocations according to their colligation and their meaning → search for appropriate semantic types
- Constant evaluation and introspection: selection of typical cases, examples, ect.
- Compiling the entry

References

Brugman, Claudia/Lakoff, George (1988): 'Cognitive topology and lexical networks', in Small, Steven L./Cottrell, Garrison W./Tenenhaus, Michael K. (eds.), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology and artificial intelligence*. San Mateo (CA): Morgan Kaufmann Publishers, pp. 477–508.

DiMuccio-Failla (forthcoming): 'A theory for usage-based cognitive lexicography'.

DiMuccio-Failla, Paolo V./Giacomini, Laura (2017a): 'Designing a learner's dictionary with phraseological disambiguators', in Mitkov, Ruslan (ed.), *Computational and Corpus-Based Phraseology, Proceedings of the second international conference EUROPHRAS 2017 (London, UK).* Cham: Springer, pp. 290–305.

DiMuccio-Failla, Paolo V./Giacomini, Laura (2017b): 'Designing a learner's dictionary based on Sinclair's lexical units by means of corpus pattern analysis and the Sketch Engine', in Kosem, Iztok/Tiberius, Carole/Jakubíček, Miloš/Kallas, Jelena/Krek, Simon/Baisa, Vít (eds.), *Electronic lexicography in the 21st century, Proceedings of the eLex 2017 conference (Leiden, Netherlands)*. Brno: Lexical Computing CZ, pp. 437–457.

DiMuccio-Failla, Paolo V./Giacomini, Laura (2022): 'A proposed microstructure for a new kind of active learner's dictionary', *Lexicographica* 38(1): 475–499.

Hanks, Patrick (2013): *Lexical analysis – Norms and exploitations*. Cambridge (MA)/London: MIT Press.

Johnson, Mark (1992 [1987]): *The body in the mind – The bodily basis of meaning, imagination, and reasoning*. Chicago/London: The University of Chicago Press.

Kilgarriff, Adam/Rychlý, Pavel/Smrž, Pavel/Tugwell, David (2004): 'The Sketch Engine', in Williams, Geoffrey/Vessier, Sandra (eds.), *Proceedings of the 11th EURALEX international congress (Lorient, France).* Lorient: UBS, pp. 105–116.

Kilgarriff, Adam/Baisa, Vít/Bušta, Jan/Jakubíček, Miloš/Kovář, Vojtěch/Michelfeit, Jan/Rychlý, Pavel/Suchomel, Vít (2014): 'The Sketch Engine: ten years on', *Lexicography* 1: 7–36.

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

Sinclair, John (2004): *Trust the text – Language, corpus and discourse*. London/New York: Routledge.