PHRASALEX II
Phraseological Approaches to Learner's Lexicography
July 22-23, 2021
Institute for Information Science and Natural Language Processing

Conceptologies: towards a phraseological framework for conceptual knowledge representation

Paolo DiMuccio-Failla
(paolodimuccio@gmail.com)

Outline

I. Wierzbicka's universal language

II. Our approach

III. The PhraseBase project

I. Wierzbicka‘s universal language

- In the historical debate between rationalists and empiricists, rationalists have asserted that we must have some innate concepts, like, for example, UNITY, PLURALITY, CAUSALITY, and POSSIBILITY.
- Leibniz went further and postulated that all human thought is constituted by combinations of such simplest concepts (cf. Wierzbicka 2011). Since elementary concepts are self-explanatory, they give us a foundation for all intelligible explanations.
- Leibniz was convinced that the task of identifying these absolutely necessary concepts, is extremely important to avoid vicious circles in the explanation of concepts: all explanations have to stop somewhere, and they can be understood only to the extent to which their primitive concepts can be understood.
- Motivated by such reflections, Anna Wierzbicka set out to find out what concepts might be common to all human languages, even the simplest ones. Her cross-linguistic investigations have led her to believe that the mental lexicon of **universal elementary concepts** is included in the common part of the lexicons of all natural languages.
- Furthermore, the common part of the grammatical systems of all natural languages constitutes a **universal mental grammar** of rules for combining concepts and make up **the language of all human thought**, which Leibniz called “*lingua naturae*”.

Example

- For example, one can find in any language not only an exact semantic equivalent of the word "*say*" but also combinations of words such as "*someone said something*," "*someone said something about something*," "*someone said something to someone else*", and "*someone said something like this [with a quotation].*"
- This means that "*say*" (in its primary sense) has a universal system of valencies and a universal combinatory potential. It is worth noting here that, contrary to what might be expected, the combination "*to say that... "* (e.g., "*he said that he would come*") is not included in this system, because not all languages have indirect speech.

(all quoted directly from Wierzbicka 2011)

Wierzbicka's conclusions

- To be clear, according to Wierzbicka, "*inside*" all languages we can find a small shared lexicon and a small shared grammar. Together, this panhuman lexicon and the panhuman grammar linked with it represent a minilanguage, apparently shared by the whole of humankind, corresponding to what Leibniz called "*lingua naturae*".
- Wierzbicka and her colleagues have been investigating this language for many years, and call it NSM, i.e. "***Natural Semantic Metalanguage***".
- This name reflects their conviction that this language provides them with a neutral tool for describing all languages.

Category	Primes
Substantives	I, YOU, SOMEONE, PEOPLE, SOMETHING/THING, BODY
Relational Substantives	KIND, PART
Determiners	THIS, THE SAME, OTHER~ELSE~ANOTHER
Quantifiers	ONE, TWO, SOME, ALL, MUCH/MANY, LITTLE/FEW
Evaluators	GOOD, BAD
Descriptors	BIG, SMALL
Mental predicates	THINK, KNOW, WANT, DON'T WANT, FEEL, SEE, HEAR
Speech	SAY, WORDS, TRUE
Actions, Events, Movement	DO, HAPPEN, MOVE
Existence, Possession	BE (SOMEWHERE), THERE IS, BE (SOMEONE/SOMETHING), (IS) MINE
Life and Death	LIVE, DIE
Time	WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT
Space	WHERE/PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE, TOUCH (CONTACT)
Logical Concepts	NOT, MAYBE, CAN, BECAUSE, IF
Intensifier, Augmentor	VERY, MORE
Similarity	LIKE/AS/WAY

(from Wikipedia: Natural Syntactic Metalanguage)

The definition of God (!)

- Who is God?

  There is someone not like people. There is no one else like this someone. God is this someone. People can't see God. At the same time, they can know God. God wants this. God is someone good. When God wants something, this something is something good. When God says something, this something is true. If God wants something to happen, it can happen because of this. God knows everything. God knows all people.

  God wants to do good things for all people. God feels something good towards all people. At the same time, God wants people to want to do good things for other people. God wants people to feel something good towards other people. All people can live with God, God wants it. At the same time, God wants all people to want it. It is very good for people to live with God. Nothing else is like this.

Criticism

- The definition of God in NSM is uncannily adequate and would probably be used by any theologian with a very limited mental lexicon. On the other hand, the following definition of "legs" is unnatural and clumsy:

  "*Legs are two parts of someone's body. These two parts are below all the other parts of the body. These two parts are long. These two parts of someone's body can move as this someone wants. Because people's bodies have these two parts, people can move in many places as they want.*"
- The truth is that we learn what legs are through our direct experience of them (seeing them in others and feeling them in ourselves), not through an convoluted rationalistic reduction to innate concepts. Wierzbicka's definition of legs looks more like a rhetorical exercise than a real explanation.

II. Our approach (1)

- Our approach is cognitivistic in nature. We are of the opinion that many natural concepts, like BIRD, are based on prototypical mental models and on radial cognitive classes (called categories by cognitive linguists – see Lewandowska-Tomaszczyk 2007). A modern electronic learner‘s dictionary could for example present such models as images or virtual 3D objects.
- Notice that different cultures may have different prototypes and cognitive classes for birds, and therefore different concepts.
- We also adhere to Fillmore‘s theory of semantic frames: the meaning of many words is not fully understood without knowledge of the context in which their referents play a role.

Our approach (2)

- Nonetheless, we think that logically complex concepts are acquired in a *leibnizian* manner, as described by Wierzbicka, simplified through the use of intermediate concepts (which may be phrasal in nature and may depend on semantic restrictions) and through language-specific grammatical constructions.
- We notice that such reduction of human thought to a finite number of both innate and experiential concepts is actually strictly limited to CONVENTIONAL, TYPICAL, NON EXPLOITATIVE (see Hanks 2013), NON CREATIVE (e.g. through original figurative language) human utterances.
- Wierzbicka has shown that the 2000 words of LDOCE's controlled defining vocabulary can be defined in a non-circular manner by means of her NSM (hence showing that all words in LDOCE are so definable). We hope, for a start, to be able to simplify her definitions and make them more natural for a learner.

Our phrasal ontology

- Semantic restrictions form a natural language ontology (see Moltmann 2016) with a hierarchy essential for understanding inheritance. For example, consider the following sentence: "*I found myself in a difficult relationship, and therefore in a difficult condition.*"
- What is the ontological connection between the concept of „relation" and that of „condition" so as to explain why they are both treated as places (in which a person finds herself)? Are they both conceived of as metaphorical places (see Lakoff & Johnson 1980)? Do their names inherit some constructions in which the word „place" participates?
- We can try and start building our ontology with the LU's denoting entities (basic nominal phrases) and add, when necessary, nominal (complex) constructions based on them, and then see, in time, weather we need even more (not naturally expressible) concepts.
- This ontology will be a sort of Wordnet 2.0, but phrasal and cognitivistic (e.g. containing cognitive metaphores and metonyms, and also nominal infinitive clauses) in nature. Therefore, calling it an „ontology" is probably not appropriate, and from now on we will use the term „conceptology" instead.

Abstractions

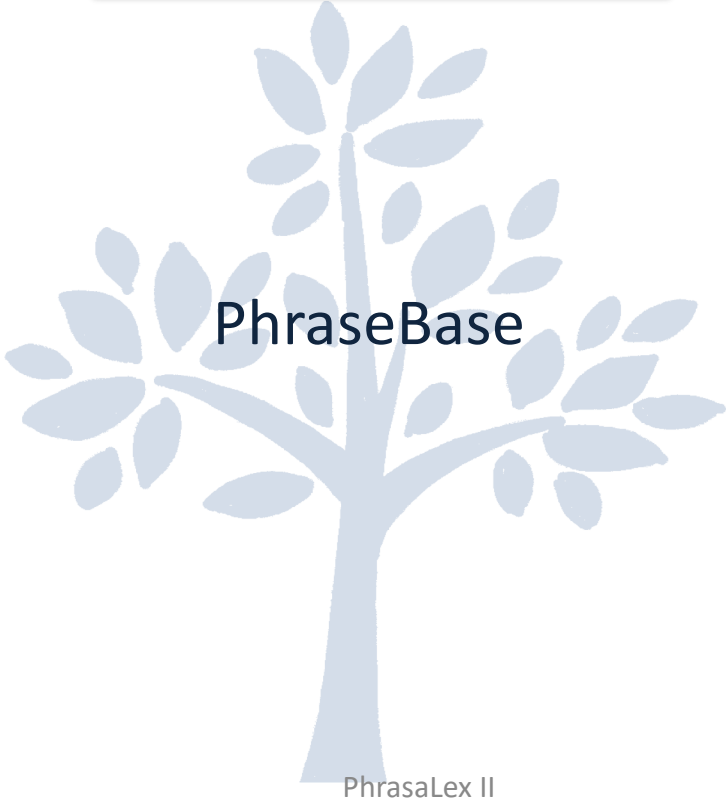
- A problem arises as to how human beings acquire “semi-abstract” concepts like EVENT, SITUATION, ACTION, ACTIVITY, and fully abstract concepts like PROPERTY, STATE, RELATION. Also, for our lexicographic purposes, how should we define such concepts? Notice that the most general of them cannot be defined by Aristotelian definitions, and are therefore often part of some circular scheme in dictionaries.
- Let us remember that people (at least non-lexicographers) often define concepts in the improper “when”-style: “The taking of the Bastille was when the Bastille was taken.” We think that this definition is grammatically incorrect but perfectly adequate.
- So we are convinced that “The storming of the Bastille took place on the 14th of July 1789.” can be defined by "The Bastille was taken on the 14th of July 1789.“
- In the case of ATTRIBUTES and RELATIONS such definitions are the only ones possible whereas in the case of EVENTS and SITUATION ostensive definitions are conceivable.

III. The PhraseBase project

- We adhere to the thesis that a language is coded by a mental lexicon of typical usage, phrasal in nature.
- We also agree that general grammatical constructions are learned together with lexical units and are in fact generalizations over sets of LU‘s.
- We put forth the thesis that a rough conceptology is also learned contextually to give an account of lexical restrictions.

Phrase-based Active Dictionary (PAD): inventory of words, lexical units and their meanings

Constructicon: structured inventory of grammatical constructions



Conceptology: hierarchy of concepts
Frame semantics: frame structures

- In the acquisition of a language, to define new words by means of already defined words and lexical restrictions (and also to define abstract words) one must have already learned basic constructions. So how does a natural path of language acquisition proceed on the structure of lexical units and general grammatical constructions? How does the tree of language develop?
- Can we mimic this growth process in an electronic resource? That is, can we ideally teach a language exactly how it is learned in nature?
- We are trying to explore the **theoretical soundness** of a non-circular PAD model extended by a structure of grammatical constructions and expanded through frame semantics and an appropriate conceptology, in an effort to capture all linguistic information necessary to write a language like a native speaker.

References

- Adamska-Sałaciak, A. (2012). Dictionary definitions: Problems and solutions. Studia Linguistica Universitatis Iagellonicae Cracoviensis, 129(4).
- Basic Formal Ontology (BFO), https://basic-formal-ontology.org/ (25/05/2021).
- Borgo, S., & Masolo, C. (2009). Foundational choices in DOLCE. In Staab, S. & Studer, R. (eds.), Handbook on ontologies. Springer: Berlin, Heidelberg, 361-381.
- Havasi, C., Speer, R., & Alonso, J. (2007). ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge. In Recent advances in natural language processing. Philadelphia, PA: John Benjamins, 261-267.
- Jezek, E. & Hanks, P. (2010). What lexical sets tell us about conceptual categories. In Lexis - Journal in English Lexicology 4.
- Johnson, M. (2015). Embodied understanding. Frontiers in psychology, 6:875
- Kilgarriff, A. (1997). I don't believe in word senses. Computers and the Humanities, 31(2), 91-113.
- Lakoff, G. & Johnson, M. (1980). Metaphors We Live By. Chicago: University of Chicago.
- Moltmann, F. (2016). Natural Language Ontology. In Aronoff, M. (ed.): Oxford Research Encyclopedia in Linguistics. Oxford: Oxford University Press.
- Moltmann, F. (2017). The Ontology of Natural Language. In Oxford Research Encyclopedia of Linguistics. Oxford UP, New York.
- Pease, A., Niles, I., & Li, J. (2002). The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web, Edmonton, Canada.
- Pustejovsky, J., Havasi, C., Littman, J., Rumshisky, A., & Verhagen, M. (2006). Towards a Generative Lexical Resource: The Brandeis Semantic Ontology. In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), 1702-1705.
- Wierzbicka, A. (2011). Common language of all people: The innate language of thought. Problems of information transmission, 47(4), 378-397.