Croatian Metaphor Repository

Kristina Štrkalj Despot
Institute of Croatian Language and Linguistics, Zagreb
e-mail: kdespot@ihjj.hr

1 Introduction: The State-of-the-art

Metaphor has been the constant object of scientific interest from Aristotle until modern day, during which time many different theoretical approaches have been performed, for example the *substitutive* and *comparative* approach, the theory of *interaction*, and most recently the *semantic* and *cognitive* approaches.

The cognitive approach differs most significantly from the traditional ones by stating that metaphor is not simply "a device of poetic imagination", an ornament or decoration in language, nor is it an anomaly or deviant use of language, but that "our ordinary conceptual system, in which we both think and act is fundamentally metaphorical in nature", as assumed by George Lakoff and Mark Johnson in *Metaphors We Live By* (1980, The University of Chicago Press, Chicago and London), where this idea and a detailed examination of its underlying processes was first extensively explored. In cognitive linguistics, the essence of *conceptual metaphor* is an understanding and experiencing one kind of thing in terms of another. The conceptual metaphor theory (CMT) has become famous and widely used among philologists throughout the world. Even in its most current form – neural theory of metaphor – which is the theoretical basis for the proposed project, it has not been significantly altered in linguistic terms.

Although there have been many influential theories regarding the cognitive basis of metaphor, the most prominent among them is Lakoff's Contemporary Theory of Metaphor (Lakoff 1993) and his Neural Theory of Metaphor (Lakoff 2008), which popularized the idea of a conceptual metaphor mapping. Within the cognitive framework of a given conceptual mapping, terms pertaining to one concept or domain (the source) can be used figuratively to express some aspect of another concept or domain (the target).

Based on a number of theories of metaphor that have been proposed (Contemporary Theory of Metaphor, Lakoff 1993; the Conceptual Mapping Model, Ahrens et al. 2003; the Structure Mapping Model, Wolff and Gentner 2000), and the Attribute Categorization Hypothesis, McGlone 1996), collections of metaphors have been assembled and published for use by researchers. Some of those are: The Master Metaphor List (MML) (Lakoff, 1994), which groups linguistic metaphors together according to their conceptual mapping; the Hamburg Metaphor Database (HMD) (Eilts and Lönneker 2002) for French and German fuses, EuroWordNet synsets with the MML source and target domains for a robust source of metaphoric semantics in those languages. In recent years, the computational linguistics community has seen substantial activity in the detection of figurative language (Bogdanova 2010; Li and Sporleder 2010, Peters and Wilks 2003, Shutova 2011) one aspect of which is the identification of metaphoric expressions in text (Fass 1991, Shutova et al. 2010, Mason 2004). Much of the early work on the identification of metaphor relied upon hand-crafted world knowledge. I.e. the CorMet system (Mason 2004) determined the source and target concepts of a metaphoric expression using domain-specific selection preferences mined from Internet resources. More recent work has examined noun-verb clustering (Shutova et al. 2010) which starts from a small seed set of one word metaphors and results in clusters that represent source and target concepts connected via a metaphoric relation. These clusters are then used to annotate the metaphoricity of text. The Metaphor Interpretation, Denotation, and Acquisition System (MIDAS) (Martin 1990) employed a database of conventional metaphors that could be searched to find a match for a metaphor discovered in text. If no match was found, the metaphoric text was replaced with a more abstract equivalent (e.g. a hyperonym) and the database was searched again. If a match was found, an interpretation mapping was activated, and the novel metaphor would be added to the database for use in future encounters.

The most current state of the art in the CMT involves theoretical and practical work done so far within a multilingual

international project *MetaNet: A Multilingual Metaphor Repository*, (*International Computer Science Institute*, Berkeley, USA). Project is led by Professor Srini Narayanan, ICSI's *Artificial Intelligence Group* leader. Linguistic analysis team leader is Professor George Lakoff (linguist, cognitive scientist, a founder of CMT and cofounder of the Neural Theory of Language and Thought project at the ICSI). PI of the proposed project is a linguistic analysis team member at the *MetaNet* project. Therefore, she is very well acquainted with these insights, many of which have not been published yet.

Theoretical background of the project proposed here involves Neural Theory of Language and Thought (NTLT) and Neural Theory of Metaphor. Within this theory, thought and language are defined as neural systems, which are:

1. systems based on neural processing and not on the formal symbols manipulation
2. *embodied* systems that reflect structure of the human body and of the physical and social surroundings
3. *evolutional systems* (in the sense that a major part of neural mechanisms used in language and cognitive functions are not only specific for humans)

Within the NTLT project a system of formal annotation of metaphors, metonymies, frames and schemas as well as of grammatical and lexical constructions has been proposed. This annotation system is coherent with all the findings of NTLT and is at the same time sufficiently precise to be suitable for demanding tasks of explicit computational modeling. This system is known under the term *Embodied Construction Grammar – ECG*. Within ECG, *construction* is defined as link between *form* and *meaning*. *Meaning* part of the construction is either directly or metaphorically linked to the *experience* and may be decomposed into embodied image schemas. ECG serves as a tool for linguistic analysis, as a tool for defining shared grammar and conceptual conventions of the linguistic community, as a computational specification for implementation of linguistic theories, as a representation of a model and theory of language acquisition, as a *front-end* system for applied tasks of language understanding, and as a high level functional description for biological and behavioral experiments.

2 Objectives and Methodology

The *Croatian Metaphor Repository* project will involve:

– theoretical cognitive linguistic research of conceptual metaphor, metonymy, image schemas and frames in the Croatian language;

– computational linguistic research on metaphor in Natural Language Processing and Artificial Intelligence involving developing tools for Automatic Metaphor Detection, Identifying Metaphorical Word Use and Automatic Extraction of Linguistic Metaphors;

– psychological experiments with purpose to further describe and explain the nature of the links between concept meaning and perception as manifested through primary metaphors.

The applied goal of the project is to create a metaphorical language database – *Croatian Metaphor Repository*.

Theoretical background for the linguistic analysis of conceptual metaphors is a Neural Theory of Language and Thought and Neural Theory of Metaphor. Formal annotation tools are taken from the embodied construction grammar. Project is methodologically coherent with the *MetaNet* Berkeley project, mentioned before.

The applied goal of the project is to create a metaphorical language database – *Croatian Metaphor Repository*.

Theoretical background for the cognitive linguistic analysis of conceptual metaphors is a Neural Theory of Language and Thought and Neural Theory of Metaphor. Formal annotation tools are taken from the embodied construction grammar. Project is methodologically coherent with the *MetaNet* Berkeley project, mentioned before.

Database will be available on the Internet. *Croatian Metaphor Repository* will include the following: a list of conceptual metaphors, a list of linguistics metaphors, and a list of image schemas, cogs and frames. Metaphors will be divided in these metaphor families: Event Structure Metaphors, Emotion Metaphors, Cascade Metaphors, Mind Metaphors, Morality Metaphors, Scalar Metaphors, Time Metaphors, Economics Metaphors, Governance Metaphors, and Well-Being Metaphors.

In addition to being differentiated by families, metaphors will be divided and searchable by *type* (primary, complex

and entailed metaphors) and by *level* (general, specific). For each metaphor type, family, level, source schema, target schema, mappings, relations, types of relations, entailed metaphors and examples of linguistic metaphors and lexical units that reflect conceptual metaphor are defined. Relations among metaphors are automatically graphically presented.

For each image schema: type, family, roles, relations, metaphors in which it serves as a source or target, inferences and lexical units from related frames are defined. Relations among schemas are also automatically graphically presented. When entering roles into schema database their relations to FrameNet roles and LU's (if applicable) are defined. Relations among metaphors include following relations: is both a source and target subcase of; is a source subcase of; is a target subcase of; is a mapping within; makes use of; has as transitive subpart 1; has as transitive subpart 2; is in some source relation to; is in some target relation to).

Where applicable (usually with primary metaphors), descriptions and/or links to the psychological experiments that has been done worldwide for a certain metaphor will be provided. It is planned within this project to conduct a series of original psychological experiments (in the collaboration with the cognitive psychologist, team member). The results of the experiments will be separately published as research papers and will also be linked to the *Croatian Metaphor Repository* Wiki page. Some experiments will involve giving examinees a task of building artefacts with a metaphorical target schema as a topic word, to see if the artefact reflects expected (or unexpected) conceptual metaphors.

2.1 Data acquisition and processing

Research corpus consists of: Croatian Linguistic Repository (Institute of the Croatian Language and Linguistics), Google examples (cautiously used), Dictionary of the Croatian or Serbian Language (known as Academy's Dictionary) and the corpus of texts for the Old-Croatian Dictionary (Institute of the Croatian Language and Linguistics). Computational linguistic research on metaphor in Natural Language Processing and Artificial Intelligence will involve Building Croatian Metaphor Repository framework (M4fw),based on web2py tools (http://www.web2py.com/); developing tools for automatic metaphor detection, identifying metaphorical word use and automatic extraction of linguistic metaphors. There are two types of computational processing within Croatian Metaphor Repository framework (M4fw): manual and automatic approach.

Within manual approach, templates are defined in which metaphors are entered:

a) Source schema
b) Target schema
c) Relations

Since Croatian language is still missing lexical repositories such as FrameNet and TreeBank, schema creation will at the same time be LU creation within Croatian version of CroLLOD (reflecting GOLD/POWLA ontology), which will be a double benefit of the project. Schemas and LU's will be stored in the Triplestore database in the N3/Turtle format, which is more recent and better approach than the one in the traditional (SQL) database, because it enables direct inclusion of data into ontologies (+ SparQl search) and the inclusion of the linked data into a cloud computing. It is also possible to employ latent semantic analysis (LSA) (Kintsch 2000, Landauer and Dumais 1997) to represent the semantic space of the metaphors in a reduced dimensionality (i.e. using SVD-singular value decomposition). We believe that the ontological knowledge encoded in the semantic relationships of metaphors represents an improvement over the distributional relatedness encoded within an LSA vector.

Croatian Metaphor Repository framework (M4fw) will include connections among newly defined metaphors with already created schemas or imported lexical units (LU's). Metaphor creation procedure will involve:

1. defining metaphor family
2. defining domain (target concept)
3. for the selected domain select schemas/LU's from the existing ones or creating new schemas
4. repeat the procedure for both concepts (source i target)
5. defining mappings
6. defining source-target relations

7. giving examples of linguistic metaphors for the conceptual metaphor that is being created
8. generating graph of metaphor and schema relations (with selected nodes and branches)
9. enlarging metaphors and schema list/index
10. storing created metaphor in the triple store
11. enabling PI and the Linguistic Adviser verification of entered metaphors

M4fw will be web application installed on a special web-server to enable work of all team members and technicians by using only web-browser (Mozzila Firefox ili Google Chrome). Each step in the metaphor creation procedure within M4fw will be supported by corresponding template for data entering or change of data.

Within the proposed project the (semi)automatic extraction of linguistic metaphors from texts is also envisaged. In a part, it will be based on the templates that have been previously manually entered. Computational analysis will enable an extraction of morphosyntactic patterns for a certain domain, which will lead to detecting new metaphors. Several approaches to automatic detection of metaphors have been proposed (Gedigian et al. 2006, Krishnakumaran and Zhu 2007; Shutova et al. 2010), all of which rely on the availability of extensive manually crafted lexical resources. Unfortunately, such resources exist only for a few resource-rich languages such as English. In this part of the research, algorithms will be developed and programmes will be created based on the merging two methods:

a) *metaphoric potential of words and word classes* (The *metaphoric potential* of words and word classes is the likelihood that the word or word class will be used metaphorically. By *metaphorical use*, we mean the use of a word to convey ideas that are not part of its basic or standard meaning. Many factors influence a word's metaphoric potential, including its conventionality as a prior metaphoric source, its familiarity, and whether it belongs to a conceptual system whose other members are often used metaphorically. We will focus here on a relatively unexplored factor: the relationality of the word's meaning (see Gentner & Kurtz 2005; Goldwater, Markman & Stilwell 2011, and Markman & Stilwell 2001)
b) *domain-sensitive semantic signature* (A domain-sensitive semantic signature is a set of highly related and interlinked senses drawn and augmented from a text that may be used to place the text within the semantic space of a metaphoric concept. We will employ a suite of binary classifiers to detect metaphoricity within a text by comparing its semantic signature to a set of known metaphors. If the semantic signature of the text closely matches the signature of a known metaphor, we will propose that it is likely to represent an instance of the same conceptual metaphor. A collection of manually detected metaphors within a particular target domain will facilitate this work. In order for an automated system to correctly understand the intended meaning of these sentences, it must first be aware that the text under consideration is not to be taken literally, and given this knowledge, it must employ all available knowledge of the underlying conceptual mapping to appropriately interpret the text in context (Mohle, Bracewell Hinote and Tomlinson 2013)

3 Afterlife

Once the structure of the database is firm and richly filled with data for a relevant number of metaphor families, which is expected to be realized within 24 months, it is planned to include examples of conceptual metaphors from other modalities too (discourse, gestures, audio-visual data, multimodal metaphors etc.). The proposed framework (M4fw) allows doing so in an appealing and elegant way. This might allow the project not to be limited by CMT, but to account for the dynamics and complexity of metaphor in action as well.

4. References

Ahrens, K. S.F. Chung, and C. Huang. 2003. Conceptual metaphors: Ontology-based representation and corpora driven

mapping principles. In *Proceedings of the ACL 2003 workshop on Lexicon and figurative language-Volume 14*, 36–42. Association for Computational Linguistics.

Bogdanova, D. 2010. A framework for figurative language detection based on sense differentiation. In Proceedings of the ACL 2010 Student Research Workshop, pages 67–72. Association for Computational Linguistics.

Bracewell, D., M. Tomlinson, and M. Mohler. 2013. Determining the conceptual space of metaphoric expressions. In Computational Linguistics and Intelligent Text Processing, pages 487–500. Springer.

Brooks, R. (2002). *Flesh and Machines. How robots will change us*. New York: Pantheon books.

Chang, N. (2009). *Constructing Grammar: A Computational Model of the Emergence of Early Constructions*. Doctoral dissertation, UC Berkeley.

Damasio, A. (1989). The Brain Binds Entities and Events by Multiregional Activation from Convergence Zones. *Neural Computation* 1 (1): 123–132.

Dodge, E. (2010). *Conceptual and constructional composition*. Doctoral dissertation, UC Berkeley.

Eilts, C. and B. Lönneker (2002). The Hamburg Metaphor Database.

Fass, D. (1991). met*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.

Fauconnier, G. i M. Turner (2003). *The Way We Think*. New York: Basic Books.

Feldman, J. (2006). *From Molecule to Metaphor. A Neural Theory of Language*. MIT Press, A Bradford Book.

Feldman, J., E. Dodge i J. Bryant (2009). A Neural Theory of Language and Embodied Construction Grammar. U: B. Heine i H. Narrog (ur.). *The Oxford Handbook of Linguistic Analysis*, 111–138. Oxford University Press.

Feldman, J. i S. Narayanan (2011). Simulation Semantics, Embodied Construction Grammar, and the Language of Events (presentation *http://www.poeticon.eu/aaai-workshop/feldman-aaai2011*)

Fillmore, C. (1982). Frame semantics. *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co.

Grady, J. (1997). *Foundations of Meaning*. Doctoral dissertation, UC Berkely.

Gibbs R. W. Jr (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge: Cambridge University Press.

Hampe, B. i J. E. Grady, ur. (2005). *From Perception to Meaning. Image Schemas in Cognitive Linguistics*. Berlin, New York: Mouton de Gruyter.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. Psychonomic Bulletin & Review, 7(2):257–266.

Klosa, Annette (2013) The lexicographical process (with special focus on online dictionaries). In: Gouws, Rufus H./Heid, Ulrich/Schweickard, Wolfgang/Wiegand, Herberst Ernst (Hgg.): Dictionaries. An international Encyclopedia of Lexicography. Supplement Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin, Boston: de Gruyter, S. 517-524. (Handbücher zur Sprach- und Kommunikationswissenschaft; 5.4).

Kövecses, Z. (2011). Recent Development in Metaphor Theory: Are the New Views Rival Ones? *Annual Review of Cognitive Linguistics* 9 (1): 11–25.

Krishnakumaran, S. and X. Zhu (2007). Hunting elusive metaphors using lexical resources. In Proceedings of the Workshop on Computational approaches to Figurative Language, pages 13–20.

Lakoff, G. i M. Johnson (1980). *Metaphors We Live By*. University of Chicago Press. [Updated version, 2002]

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. University of Chicago Press.

Lakoff, G. (1993). The contemporary theory of metaphor. U: A. Ortony (ur.). *Metaphor and Thought*. New York: Cambridge University Press.

Lakoff, G. (1994). Master metaphor list. University of California.

Lakoff, G. i M. Johnson (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books.

Lakoff, G. (2008). The neural theory of metaphor. U: R. Gibbs (ur.). *The Cambridge Handbook of Metaphor and Thought*. New York: Cambridge University Press.

Lakoff, G. (2012). Explaining Embodied Cognition Results. *Topics in Cognitive Science* 4 (4): 773–785.

Li, L. and C. Sporleder (2010). Using gaussian mixture models to detect figurative language in context. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 297–300. Association for Computational Linguistics.

Martin, J.H. (1990). A computational model of metaphor interpretation. Academic Press Professional, Inc. Mason, Z.J. (2004). CorMet: A computational, corpusbased conventional metaphor extraction system. Computational Linguistics, 30(1):23–44.

Markman, A. B., & Stilwell, C. H. (2001). Rolegoverned categories. *Journal of Experimental & Theoretical Intelligence, 13*, 329-358.

McGlone, M. S. (2001). Concepts as metaphors. U: S. Glucksberg (ur.). *Understanding figurative language: From metaphors to idioms*, 90–107. Oxford, UK: Oxford University Press.

Mohler, M., D. Bracewell, D. Hinote and M. Tomlinson (2013). Semantic Signatures for Example-Based Linguistic Metaphor Detection. U: *Proceedings of the First Workshop on Metaphor in NLP*, Atlanta, Georgia, 13 June 2013: 27–35, Association for Computational Linguistics.

Murphy, G. L. (1996). On metaphoric representations. *Cognition* 60: 173–204.

Narayanan, S. (1997a). *KARMA: Knowledge-based Action Representations for Metaphor and Aspect*. Doktorska disertacija, UC Berkeley.

Narayanan, S. (1997b). Talking the talk is like walking the walk: A computational model of verbal aspect. *Proceedings of the 19th Cognitive Science Society Conference*.

Peirsman, Y.; D. Geeraerts (2006). Metonymy as a prototypical category. *Cognitive Linguistics* 17 (3): 269–316.

W. Peters and Y. Wilks. 2003. Data-driven detection of figurative language use in electronic language resources. Metaphor and Symbol, 18(3):161–173.

Pragglejaz Group (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol* 22 (1): 1–39.

Shutova, E. 2010. Models of metaphor in NLP. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 688–697. Association for Computational Linguistics.

Shutova, E. 2011. Computational approaches to figurative language. Ph.D. thesis, University of Cambridge.

Sinha, C. i K. J. de López (2000). Language, Culture, and the Embodiment of Spatial Cognition. *Cognitive Linguistics* 11 (1/2): 17–41.

Steen, G. (1999a). Metaphor and discourse: Towards a linguistic checklist for metaphor analysis. U: L. Cameron i G. Low (ur.). *Researching and Applying Metaphor*, 81–104. Cambridge: Cambridge University Press.

Steen, G. (2008). The Paradox of Metaphor: Why We Need a Three-dimensional Model of Metaphor. *Metaphor and Symbol* 23 (4): 213–241.

Steen, G. (2008a). *Finding Metaphor in Grammar and Usage: A Methodological Analysis of Theory and Research*. Amsterdam: John Benjamins.

Wolff, P. and D. Gentner. 2000. Evidence for role-neutral initial processing of metaphors. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(2):529.