Embodied Language Processing: a new generation of language technology

K. Pastra, E. Balta, P. Dimitrakis, G. Karakatsiotis

Cognitive Systems Research Institute, 7 M. Prantouna Street, 11525, Athens, Greece

{kpastra, ebalta, pdimitrakis, gkarakatsiotis}@csri.gr

Abstract

At a computational level, language processing tasks are traditionally processed in a language-only space/context, isolated from perception and action. However, at a cognitive level, language processing has been shown experimentally to be embodied, i.e. to inform and be informed by perception and action. In this paper, we argue that embodied cognition dictates the development of a new generation of language processing tools that bridge the gap between the symbolic and the sensorimotor representation spaces. We describe the tasks and challenges such tools need to address and provide an overview of the first such suite of processing tools developed in the framework of the POETICON project.

Introduction

Natural language processing tools have a history that can be traced back to the very early days of Artificial Intelligence; a number of language related tasks have been dealt with rule-based, statistical, connectionism or other approaches and have reached a level of maturity that allows for their use by laymen. For example, Statistical Machine Translation has reached unprecedented performance during the last decades boosting research in this field and leading to adoption of the related technology for general use (cf. for example the Google Translation service). Similarly, automatic text indexing and retrieval technology is behind powerful search engines on the web, while dialogue systems are increasingly used in interactive voice response applications as well as in human-robot communication.

However, language technology has in many cases reached its limitations; hardly can new systems advance performance considerably, while scalability beyond domain-specific applications for everyday use remains a quest. A reason for this may have been the fact that language processing has been largely confined within the language space; in other words, language analysis and generation have been predominantly treated as language-only processes, isolated from perception and action.

However, recent advances in neuroscience and cognitive psychology point to the opposite direction; language has been shown to be tightly connected to perception and action. Embodied cognition research advocates the embodied nature of language and therefore suggests that language processing should not take place in a cognitive vacuum.

In this paper, we argue that a new generation of language technology emerges out of the embodied cognition view of language, one that has to address the challenge of bridging the gap between symbolic representations and sensorimotor ones. In what follows, we review the related evidence from embodied cognition research, we sketch the tasks and challenges such language tools are required to address and provide an overview of the first suite of embodied language processing tools, which we have developed within the European funded project POETICON (www.poeticon.eu). The tools have been integrated in a humanoid platform for performing everyday tasks.

Language and Embodied Cognition

Recent years have seen an increasing body of experimental evidence suggesting a tight relation between language and action. Part of this evidence sheds light on the role of the (visuo)motor system in language comprehension. For example, the contribution of motor circuits in the comprehension of phonemes, semantic categories and grammar has been reported in both neuroimaging studies (Pulvermuller 2005) and ones involving transcranial magnetic stimulation and patients with certain types of brain lesions (Fadiga et al. 2009, Pulvermuller and Fadiga 2010). In cognitive experiments, language understanding has been found to activate motor simulation (Gallese andLakoff 2005, Glenberg and Robertson 2000, Zwann and Teylor 2006, Fischer and Zwann 2008, Glenberg 2008), i.e. it has been found to recruit the same sensorimotor areas that are activated when interacting with objects or simulating the state of the world denoted verbally (Zwann 2004).

It has been claimed that sentences are understood by the simulation of actions that underlie them (Glenberg and Robertson 2000, Feldman 2008). In particular, Glenberg's Indexical Hypothesis claims that when reading a sentence, first words and phrases are indexed to the denoted objects in the environment, then the affordances of such objects are derived and then these affordances are meshed into a coherent set of actions; language is used to guide motor control to produce an effective action or simulation. Abstract concepts (Barsalou 2008a, Glenberg 2008) and metaphoric language (Lakoff and Johnson 1980, Feldman 2008, Feldman 2010) have also been shown to be grounded in the sensorimotor system too (abstract source concepts grounded to concrete target domains). In the same line of language creativity research, the Conceptual Blending theory (Fauconnier and Turner 2002) explains creativity as a semantic process that operates on the output of perception and interaction with the world to create new structures (blends).

Indeed, embodied cognition has imbued language analysis theories and has boosted research in cognitive linguistics. However, formalisation of such approaches algorithmically is still in its infancy. It is mainly research on construction grammar that leads in this research front, i.e. a number of formalisms which consider language syntactic patterns (constructions) to have their own semantics (i.e. language syntax to be semantically motivated) and advocate a syntax-lexicon continuum (cf. for example, Bergen and Chang in press, Steels and Beule 2006, Chang 2008, Goldberg 2009, Goldberg 2010). Currently, only one such formalism is fully operational (Steels and Beule 2006) and is being used in (artificial) language acquisition experiments with artificial agents. No natural language processing tools exist that process natural language according to embodied principles, so that it is associated directly with the sensorimotor domain.

A new generation of language processing tools is needed for analysing natural language so that:

(a) *it is connected to what it denotes in the sensorimotor world (including abstract concepts),* and

(b) *it infers common sense information that is not explicitly expressed, including implied concepts and semantic relations*.

Such tools should be able to process, for example, action denoting verbal input and both extract and infer the corresponding action-related sensorimotor information that is needed by an agent to perform the action. The minimum requirement for such tools should be to extract and infer basic sensorimotor action constructions; we refer to constructions that experimental evidence has shown humans to use for expressing action related information regardless word order in their native languages (Goldin-Meadow et al. 2008). In particular, these experiments have shown that people follow an Actor-Patient-Act order when describing events through gestures or when reconstructing them through pictures; the same pattern goes beyond communication needs and beyond word order in people's native languages. This pattern corresponds to the Subject-Object-Verb syntactic pattern, or -in sensorimotor domain terms (and following the minimalist grammar of action, Pastra and Aloimonos, in press)- it corresponds to "Tool-Affected Object-Action Terminal" patterns.

Embodied Language Processing Tasks

The new generation of language processing tools needs to go beyond current technology that segments text into tokens, attributes part of speech tags to tokens, and performs morphological, syntactic analysis and lexical semantic analysis. Actually, these tasks need to be reformulated to address the embodied cognition challenges. We have identified four such tasks, as follows:

Concept segmentation: this is a task of automatically segmenting language into units (word(s), phrases) that refer to unique **embodied** rather than lexical concepts, i.e. they have a unique reference (entity, movement, descriptive feature or abstract concept). The task requires one to go beyond the traditional [word : meaning] or [multiword expression : meaning] associations and incorporate cases of whole e.g. verb phrases corresponding to a single *meaning*. For example, '*to slice the tomato with the knife*' is a specific movement defined by the use of a specific instrument on a specific object in a particular way. This is a *single* embodied concept denoting a movement which is inherently related to a specific tool and a specific affected object. An embodied language tokenizer needs to go beyond word boundaries, beyond lexical concepts, to verbal phrases that correspond to embodied concepts.

Concept type tagging: the task involves automatic attribution of concept type to every embodied concept identified in the previous task. It goes well beyond part of speech tagging (e.g. noun, verb, adjective tag attributions), to attribution of an entity, movement, descriptive feature or abstract concept type. For example, *'the walk', 'walking', 'to walk'*, are all language representations of the very same embodied concept, that of a specific body movement. The embodied language tagger must be able to go beyond language specificity to the denoted meaning. It should also be able to distinguish abstract concepts into ones with concrete origin (i.e. categorization abstraction, e.g. 'furniture' is an abstract concept of entity origin related to a number of concrete entities such as 'sofa', 'dining table' etc.) and ones of non-concrete origin (e.g. poverty). Another challenge in this task is the need for distinguishing between literal and figurative use of words (e.g. 'tiger' the animal and 'tiger' the courageous or aggressive person).

Semantic relation extraction: the task involves the automatic extraction of binary semantic relations between embodied concepts, such as 'action-tool', 'action-affected object', 'entity-location' relations and many others. The extraction of such relations requires a further attribution of semantic type to each concept that denotes the role of the concept within a specific e.g. action denoting structure. The attribution of such type goes beyond syntactic criteria and should actually deal with syntactic variability.

Common sense generation: this is a task of automatically inferring concepts (and relations) that are verbally implicit, but essential in the sensorimotor space. It is a task of making explicit common sense information that due to its nature is never/rarely expressed explicitly through language (e.g. the simple verbal request to 'touch wood' leaves information on the tool to be used for the task, implicit, i.e. the 'hand'). This implicit information

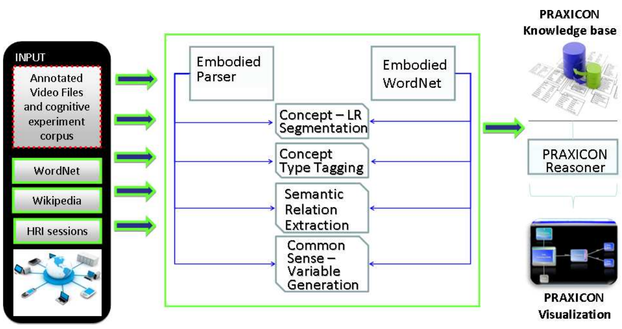

Figure 1: Embodied Language Processing Tools

may remain a variable (i.e. only its type and relation to other concepts is generated), or it may be solved by attributing one (or more, if appropriate) values to it. For example, the verbal request to 'stir the soup' does not include information on the tool to be used for stirring (the tool feature remains a variable). It could be a number of different tools, e.g. the commonly used 'spoon', or a 'stirring stick'. In some cases, such common sense information is available in general textual resources for mining and filling in the implied information. When it is not, visual object and action parsers are necessary.

Figure 1 presents these four processes as carried out by two different tools: the Embodied Language Parser and the Embodied WordNet, each working on different types of textual resources: The Embodied WordNet processes the WordNet Lexical Database (Fellbaum 1998) for extracting the related information, while the Embodied Language Parser runs on free text coming from verbal requests in human-robot interaction sessions, with the vision to be extended so that it runs on dictionary definitions, large text corpora or even Wikipedia. The information extracted is fed to an embodied concept knowledge base (e.g. the PRAXICON, cf. Pastra 2008) and the associated Reasoner.

The Embodied WordNet

The Embodied WordNet runs on the WordNet lexical database (Fellbaum 1998) to identify unique concepts, attribute concept type information and extract pragmatic relations between concepts. It aims at transforming this database from a sense-based to a reference-based resource, splitting, merging and analysing synonymy sets (synsets) using only information from within WordNet itself, i.e. by exploiting the WordNet taxonomy and word-definitions. In its current version, the tool processes the biggest part of WordNet (i.e. the noun hierarchy, the related verbs and adjectives). It actually grounds WordNet, re-processes and extends its content for populating the PRAXICON embodied concept database. It addresses a big challenge of bridging the gap between language-based conceptual categorization of the world and denoted sensorimotor references and populates the PRAXICON with more than 120.000 concepts and 217.000 conceptual relations. These concepts have been enriched with corresponding visual representations form the ImageNet database (Deng et al. 2009). In particular, the module performs the following:

(a) It tags each entry in the WordNet as being either an entity, a movement, a feature or an abstract concept, regardless of its grammatical category. This is a process of transforming the sense-based WordNet lexicon entries into reference-based ones, i.e. ones that can be directly linked (grounded) to sensorimotor representations;

(b) It distinguishes concepts into literal vs. figurative and basic-level vs. non-basic level ones. The latter distinction applies to abstract concepts only and involves an algorithmic implementation of identifying verbal concepts that express the 'basic-level of categorisation' (Rosch 1978 on the basic-level theory). Abstract concepts which refer to concrete entities, movements or features (e.g. 'cup' refers to a class/category of visually and/or functionally similar entities such as 'coffee-cup', 'chalice' etc.) are distinguished from ones that do not, such as 'poverty'. The module provides this information for each abstract concept by assigning an 'origin' attribute to each one of them (e.g. 'entity-origin'); language-based criteria (morphology) and tree network topology ones (e.g. connectedness, density and position) and are combined for performing this task.

(c) It uses the WordNet semantic relations and other pragmatic relations that it mines from the lexicon definitions to interconnect the grounded concepts formulating a grounded semantic network. Furthermore, the module applies semantic and linguistic criteria to enrich the network with metonymic and metaphoric relations not explicitly denoted within the WordNet database.

The Embodied Parser

The Embodied Parser runs on free text to extract embodied concepts and relations, including information regarding verbally expressed analogies, justifications, and comparisons; its output comprises one (or more) action tree(s) denoted in text. The tool has built on findings from a rule-based extraction of this information from manually annotated corpora; in its current version, it is limited to parsing verbal requests for tasks to be performed, of the form 'verb-object-prepositional phrase denoting tool'. It has been trained to recognise a number of diverse syntactic patterns that one may use to express such information (and/or imply common sense information), making use of VerbNet (Kipper-Schuler 2005), a language processing resource that classifies verbs according to the syntactic patterns in which they may appear.

The tool has been used mainly within the POETICON robotic demonstration for allowing an agent to understand and satisfy a verbal request; however, it is envisaged that future extensions of this module will perform concept and relation mining from large textual resources, for extending the PRAXICON embodied concept knowledge base. The main contribution of this tool in its current version is that it implements -for the first time- a mapping between traditional language syntax trees and sensorimotor action trees.

This information is extracted in the form of a high-level action syntax tree, which consists of: an *action* concept and its *inherent* relations with the *tool*(s) used to carry out the action as well as the *affected object*(s) and the *high-level goal* of the action. This type of syntax trees are not traditional language syntax trees, but actually follow the minimalist action grammar theory developed by Pastra and Aloimonos (in press). They are of the same form as these sensorimotor action trees, but instantiated at a high level, with information that is provided by language. They actually *implement a mapping between language structure and action structure*.

In other words, the module parses a verbal request in order to:

(a) identify action, action-tool, action-affected object and action-purpose concepts mentioned, and

(b) create a labeled, directed acyclic graph with this information.

In its current implementation, the module parses sentences that express requests for biological transitive (e.g. stir the coffee) actions. The challenge for the module lies in two facts:

(a) Syntactically, there are many different ways to express requests for the same action e.g. 'stir the coffee with the spoon' vs. 'could you please stir the coffee with the spoon?' vs. 'I would appreciate it if you could stir the coffee with the spoon'. In all these examples, the task to be performed remains the same, though the verbal expression ranges from a simple imperative of the form 'Verb-Noun Phrase' to more complex constructions including questions and even conditional sentences;

(b) Frequently, verbal expressions are general or vague leaving a lot of information unspecified; in particular, common sense information related to the tool(s) to be used for carrying out an action may remain implicit. For example, a request may be more or less specific: 'stir the coffee' vs. 'stir the coffee with the stirrer' vs. 'stir the coffee with the red spoon'. The semantic analyzer goes beyond variability in syntactic expression and identifies 'missing information'; for the latter, it generates variable concepts that are attributed a semantic/pragmatic role and are included in the extracted graph; the concepts are instantiated by a Reasoner to which the action tree/graph is passed.

In order to address these two challenges, the semantic analyzer was trained using VErbNet (Kipper-Schuler 2005), a syntactic pattern lexicon which groups verbs into classes according to the syntactic patterns they share. The lexicon employs Harris' distributional analysis theory, according to which words that tend to occur in the same (language) contexts tend to have similar meanings. Among others, it provides information on the variety of syntactic patterns in which the members of each verb class appear in English sentences. The syntactic patterns in VerbNet comprise of part of speech and phrase type information, lexical information, as well as basic lexical-semantic information such as 'agent', 'patient' etc. For example, one of the syntactic patterns in which members of the "carve" verb class appear in is the following:

<NP value="Agent">
<VERB>
<NP value="Patient">
<PREP value="with">
<NP value="Instrument">

An example sentence with this syntactic pattern is: "*Carol crushed the ice with a hammer*".

We selected 46 classes of transitive verbs and extracted such patterns from the corresponding examples in order to create a list of potential syntactic patterns that may appear in sentences and to map these patterns with the semantic information needed in action trees (i.e. action/purpose, tool, and affected object). We derived approximately 60 such syntactic patterns which reflect the syntactic variability of verbal expressions that the semantic analyzer can currently handle.

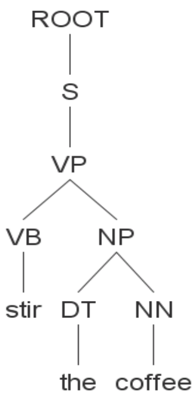

Figure 2: Syntactic Tree

Once the verbal request is submitted to the module, the corresponding sentence is parsed using the Stanford Parser (Klein and Manning 2003), which produces a language syntax tree (cf. a simple example in figure 2). The known syntactic patterns are compared against this structure for identifying the one that best "fits" to the sentence. The 'fitting' process provides a way of attributing both concept type (e.g. movement, entity etc.) and action, tool, affected object roles to parts of the sentence (corresponding largely to verb, noun phrase etc.).

For example, given the request "Stir the coffee", the Semantic Analyzer invokes the Stanford text parser which produces the linguistic syntax tree shown in figure 2. Then, the Embodied Parser serializes the tree into a lexico-syntactic pattern such as: *VBP_stir NP_the_coffee*, and applies the syntactic-semantic pattern that best fits this serialization, according to its training. The chosen pattern is: *VERB NP_Affected Object*. So, the module maps the "coffee" to an entity concept that has the role of the affected object for the action concept "stir".

Furthermore, it detects that no information on the tool used for performing the action is given and generates a variable concept that needs to be instantiated with appropriate tool-candidates. Figure 3 shows the high level action tree generated by the analyzer. A´´ denotes the whole action sequence needed for performing the requested action, while A´ are its constituents. In this stage, since we go from language to the denoted action to be

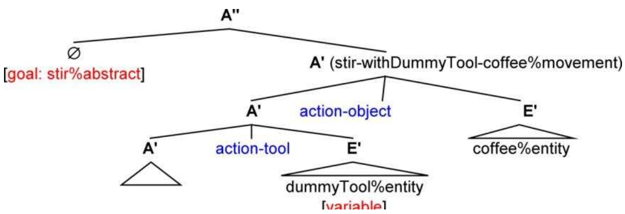

Figure 3: High level action tree

performed, no full analysis of the action constituents is given (denoted by the triangle in the lowest-level A´). However, entity-constituents of the action are given, such as the tool (a variable) and the affected object. These are inherent relations of the action. Also the goal of the whole action sequence is clearly denoted (no matter how the stirring will be done in terms of exact movement constituents, tools and affected objects, the general goal is the one denoted by language: the stirring).

Also, note that the type of the concepts that label the tree nodes as given by the module is an approximate one (it does not correspond to the concept type as known in the Praxicon knowledge base from the Embodied WordNet module, since the analyzer has no way of knowing whether the particular concepts express, indeed, very specific movements/entities or abstractions of movements/entities at basic level or beyond). The mapping to the appropriate concept type is performed at a subsequent step.

This tree is further processed by the Praxicon Reasoner for generating the fully elaborated action structure denoted by the verbal command.

The Embodied Reasoner

The PRAXICON Reasoner comprises of a number of modules that come part and parcel with the PRAXICON knowledge base and actually exploit the characteristics of each concept's semantic network topology for addressing a number of phenomena and a number of application specific reasoning needs. In particular, the PRAXICON reasoner performs the following tasks:

Simple Reasoning: The reasoner computes the shortest path that connects a list of concepts. To find this path, the module performs a Breadth-First-Search (BFS) from all the concepts that are on the list, so that if there is a path, it will be found and the first to be found will be the shortest one. The reasoner performs BFS on all endpoints of the path, so that to constrain the search space (since each tree of the BFS is relatively smaller than having just one BFS tree starting from any endpoint).

Generalisation through Variable solving: As mentioned in the previous sections, the Embodied Language Processing tools may generate a number of 'variables' as stand alone embodied concepts and/or inherent concepts that define more complex ones; these are contributed by the language representation space and actually they are key/unique contributions of language to the sensorimotor space for generalisation purposes. However, embodied cognition applications require variable resolution. To this end, our reasoner takes the following steps:

(a) It finds all Concepts that the variable is connected to;

(b) From those Concepts, it removes the ones that do not contain a relation of the same type to the one that connects each concept to the variable that is to be solved;

(c) For each of the remaining Concepts, it computes the similarity of the Concept to the variable; the higher the similarity, the better. Similarity Computation takes advantage of a number of characteristics of the concepts' semantic network topologies. It is a function of:

- The number of relations they have (connectedness)
- The concepts they are connected with;

- Their taxonomic relations; for example, when the concepts that are compared do not share any connections with other concepts, the module goes beyond exact matching of the latter, expanding the matching with other concepts that have a type-token relation with them.

- The type of relations through which they are connected; shared *inherent* relations take precedence.

Generalization through substitution: As mentioned above, the Reasoner makes use of the knowledge base to find the optimal path linking a number of concepts. This simple reasoning process forms the bases of a mechanism that suggests possible substitutions of a specific concept with other more or less similar ones for accomplishing a task. This is actually a mechanism for controlled generalisation of knowledge, since the algorithm provides scores of 'effectiveness' of substituting one concept with another, while it hinders generalizations that are farfetched. The algorithm has been used in the POETICON robotic demonstration to allow the robot to use its existing knowledge to perform a user defined task under a novel situation .

The reasoner takes advantage of taxonomic relations (isA, type-token relations) and the indication of basic level categorization in such hierarchies, to find for each concept A to be substituted:

- its '*children concepts*' (hyponyms) and 'parent concepts' (hypernyms) up to and including basic level concepts; these are listed in ascending order according to their distance from concept A;
- its '*sister concepts*' (i.e. all concepts that have the same basic level concept parent) and their children (ascending by distance);
- the '*sister concepts' of its basic level parent and their children* (ascending by distance);
- Concepts that are *of the same type* with Concept A and are related to it with relations such as: container-content (one step distance in Concept's A graph);
- Concepts that are *of different type* to Concept A and are related to it directly (e.g. descriptive feature relations etc.); inherent relations take precedence over non-inherent ones.

For example, let's assume that a request to '*stir the coffee*' is given to an Embodied Language Processing Tools - enhanced robot. By mining text corpora, the robot knows that it is spoons that are normally used for stirring coffee. However, no spoon is available in its environment. Instead, a *plate, a cup, a knife and a pencil* are available. The reasoner will help the robot to succeed in the task by using a different tool, or to be precise, by using the most suitable object in its environment for stirring, reaching creativity, but avoiding improbable reactions, i.e. mistakes. The problem amounts to substituting the 'spoon' concept with another one that will take up the same role in the specific motor program, more or less efficiently. By implementing the above mentioned criteria, the reasoner will suggest – in this example – the use of the 'knife' as the best alternative for stirring the coffee, followed by the 'pencil', and then by the 'cup' or 'plate' the latter with very low 'effectiveness' score that would actually prohibit their use.

Conclusion

The development of a new generation of language processing tools that employ the embodied cognition perspective opens up a new direction in research that may be key to (a) introducing language dynamically in the 'agent sensing', 'agent acting', 'agent learning' loop and (b) taking language processing tasks to the next level allowing for scalability and generalisation beyond domain specific applications. In this paper, we sketch the tasks and challenges for such tools, and the abilities of the very first version of an embodied language processing suite.

Acknowledgements

Research reported in this paper has been funded by the European Commission, Framework Programme 7, Grant 215843 – POETICON Project, which we gratefully acknowledge.

References

Barsalou L. (2008). *Grounding Symbolic Operations in the Brain's Modal Systems*. In Semin G. and E. Smith (eds). *Embodied Grounding: Social, Cognitive and Neuroscientific Approaches*, ch. 1, pp. 9-42, Cambridge University Press.

Bergen B. and N. Chang (in press). *Embodied Construction Grammar in Simulation-Based Language Understanding*. J.-O. Östman and M. Fried (eds.). *Construction Grammar(s): Cognitive and Cross-Language Dimensions*. John Benjamins.

Chang N. (2008). *Constructing grammar: A computational model of the emergence of early constructions*. Computer Science Division, University of California at Berkeley, PhD dissertation.

Deng J., A. Berg, K. Li and L. Fei-Fei (2010). *What does classifying more than 10,000 image categories tell us?* In Proceedings of the 12th European Conference of Computer Vision (ECCV).

Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei (2009). *ImageNet: A Large-Scale Hierarchical Image Database*. IEEE Computer Vision and Pattern Recognition (CVPR).

Fadiga L., Craighero L. and D'Ausilio A. (2009). *Broca's area in language, action, and music*. Annals of New York Academy of Science, 1169:448-58.

Fauconnier G. and M. Turner (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities.* New York: Basic Books.

Feldman J. (2008). *From Molecule to Metaphor: A Neural Theory of.* Cambridge : MIT Press.

Feldman J. (2010). *Embodied language, best-fit analysis, and formal compositionality*. Physics of Life Reviews, 7(4), pp. 385-410.

Fellbaum C. (ed.) (1998). *WordNet: An Electronic Lexical Database.* The MIT Press, Cambridge, MA.

Fischer M. and Zwann R. (2008). *Embodied language: A review of the role of the motor system in language comprehension.* The Quarterly Journal of Experimental Psychology, vol. 61(6), pp. 825–850.

Gallese V. and Lakoff G. (2005). *The brain's concepts: the role of the sensorimotor system in conceptual knowledge.* Cognitive Neuropsychology, vol. 22(3/4), pp. 455-479.

Glenberg A. (2008). *Toward the Integration of Bodily States, Language and Action*. In Semin G. and E. Smith (eds). *Embodied Grounding: Social, Cognitive and Neuroscientific Approaches.* ch. 2, pp. 43-70, Cambridge University Press.

Glenberg A. and Robertson D. (2000). *Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning*. Journal of Memory & Language, vol 43(3), pp. 379-401.

Gluck, M., Corter, J. (1987). *Information, uncertainty, and the utility of categories.* In Proceedings of the seventh annual conference of the cognitive science society, pp. 283–287.

Goldberg, A. (2010). *Verbs, Constructions and Semantic Frames.* In Rappaport Hovav, M., Doron, E. and Sichel, I. (ed.). *Syntax, Lexical Semantics and Event Structure*, Oxford University Press.

Goldberg, A. and Suttle, L. (2009). *Construction Grammar.* Interdisciplinary Reviews: Cognitive Science, vol. 1(4), pp. 1-10.

Goldin-Meadow S., So W., Özyürek A. and Mylander C. (2008). *The natural order of events: How speakers of different languages represent events nonverbally*. PNAS vol. 105:9163-9168.

Klein D. and C. Manning (2003). *Accurate Unlexicalized Parsing.* Proceedings of the 41st Meeting of the Association for Computational Linguistics, pp. 423-430.

Kipper-Schuler K. (2005). *VerbNet: a broad coverage comprehensive verb lexicon*, PhD thesis, University of Pennsylvania.

Lakoff G. and M. Johnson (1980). *Metaphors we live by.* Chicago: University of Chicago.

Pastra K. (2008). *PRAXICON: The Development of a Grounding Resource.* In Proceedings of the 4th International Workshop on Human-Computer Conversation, Bellagio, Italy.

Pastra K., Dimitrakis P., Balta E. and Karakatsiotis G. (2010) *PRAXICON and its language-related modules.* In Proceedings of Companion Volume of the 6th Hellenic Conference on Artificial Intelligence (SETN), pp. 27-32, Athens, Greece.

Pastra K. and Y. Aloimonos (in press). *The Minimalist Grammar of Action.* Philosophical Transactions of the Royal Society - London B Biological Sciences.

Pulvermüller, F. (2005). *Brain mechanisms linking language and action*. Nature Review of Neuroscience, vol. 6, pp. 576-582.

Pulvermüller F., & Fadiga, L. (2010). *Active perception: Sensorimotor circuits as a cortical basis for language*. Nature Reviews Neuroscience. Vol. 11(5), pp. 351-360.

Rosch E. (1978). *Principles of Categorization.* In E. Rosch and B. Lloyd (eds.). *Cognition and Categorization*. Ch. 2, pp. 27-48, Lawrence Erlbaum Associates.

Steels, L. and De Beule, J. (2006). *Unify and Merge in Fluid Construction Grammar.* In Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication, Berlin.

Zwaan R. (2004). *The immersed experiencer: Toward an embodied theory of language comprehension.* In Ross B. (Ed.). *The psychology of learning and motivation*. Vol. 44, pp. 35–62, New York: Academic Press.

Zwaan R. and Taylor L. (2006). *Seeing, acting, understanding: Motor resonance in language comprehension.* Journal of Experimental Psychology: General, vol.135.