

Towards a System Architecture for Integrating Cross-modal Context in Syntactic Disambiguation

Patrick McCrae and Wolfgang Menzel

CINACS Graduate Research Group
Department of Informatics, Hamburg University
Vogt-Kölln-Straße 30, 22527 Hamburg, Germany

Abstract. Most natural language utterances are inherently ambiguous, which results in semantic underspecification. Yet, despite the omnipresence of ambiguity, human communication still succeeds in most cases and even displays a remarkable robustness – quite in contrast to the majority of natural language applications today. The reason for this is that in processing an ambiguous utterance humans also integrate information from sources other than the utterance itself and thus have access to additional knowledge to enrich the semantic specification that guides disambiguation. One important such source of additional semantic knowledge in humans is sensory input from cross-modal perception.

In this paper we describe a system architecture motivated by effects during human sentence processing which permits to study the integration of cross-modal context knowledge in syntactic parsing. We hypothesise that integrating cross-modal context into syntactic constraint dependency parsing will significantly and substantially improve the accuracy of structural ambiguity resolution.

1 Introduction

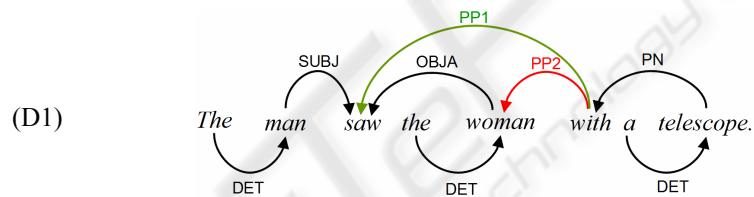
Considering the complexity of factors contributing to successful natural language interaction, human language processing is surprisingly robust against ambiguous or ungrammatical input – in fact, much more robust than the majority of present-day natural language processing (NLP) systems. An important reason for this robustness in human sentence processing is the access to information sources other than the linguistic material. In the computation of an utterance's overall meaning, humans do not only analyse the linguistic material in isolation but also incorporate additional linguistic and extra-linguistic information to establish reference and resolve various types of ambiguity. Typical sources of additional knowledge include cross-modal sensory information [1], discourse history, context [2] and common sense or world knowledge [3]. All of these help constrain utterance context and thus facilitate the computation of the most plausible sentence meaning resulting from the overall context evaluated.

One extensively studied example for syntactic ambiguity is prepositional phrase (PP) attachment. It is of particular interest because in many cases disambiguation cannot be performed on linguistic grounds alone. Consider the classic example (S1).

(S1) *The man saw the woman with a telescope.*

(S1) has at least two readings, depending on which constituent the PP *with a telescope* is considered to attach to. In the first reading, *with a telescope* modifies the act of seeing and functions as the thematic INSTRUMENT for seeing. In the second reading, the term *telescope* modifies *the woman* and thus functions as the thematic COMITATIVE expressing the concept of companionship.

Observe that from a purely linguistic point of view both readings and structural interpretations are equally valid. A conclusive attachment preference requires additional, extra-linguistic information which other sources such as context may provide. A key source of contextual information in human communication is cross-modal sensory perception [4], [5]. Given such context information – e. g. by additionally seeing the image of a man looking through the telescope – we have sufficient extra information to favour one reading over the other and hence arrive at different dependency structure representations in (D1); attachment PP1 for the thematic role INSTRUMENT and attachment PP2 for the thematic role COMITATIVE.



While a large number of NLP techniques for handling PP attachment exist, the majority of these approaches rely on lexical or syntactic properties of the input sentence or statistical distribution patterns of the preposition relative to its attachment constituent. All of these approaches are based on properties of the sentence material alone; none of them incorporate extra-sentential context information although utterance context is known to direct semantic and structural interpretation in human sentence processing [2]. If we wish to enable an NLP system to make informed context-dynamic decisions on structurally ambiguous utterances, we need to provide the basis for such decisions by making suitable representations of context information available for integration into the syntactic decision process.

Starting out from CDG, a weighted-constraint dependency syntax parser for German, our goal is to develop an architecture capable of integrating cross-modal context in syntactic parsing. The architectural integration of cross-modal context simulates the effects of the cognitive processes during human sentence processing. We hypothesise that integrating cross-modal context will significantly and substantially improve the accuracy of structural ambiguity resolution.

The structure of this paper is as follows: The *Literature Review* section outlines current approaches to PP attachment, one notoriously hard-to-handle type of structural ambiguity, and provides an overview over the NLP challenges in assigning thematic

roles. From these we derive general requirements for our Context Integration Architecture. In Section 3 we present the architecture, first conceptually and then with an implementation focus. In the *Future Work* section we give an outlook onto which cognitive and linguistic phenomena we intend to study.

2 Literature Review

Most heuristics for syntactic ambiguity resolution in broad-coverage NLP applications today are based on the syntactic or syntacto-lexical properties of the sentence material rather than contextual semantics. These heuristics derive their answer to a syntactic problem from a primarily syntax-oriented procedure. Given that accurate thematic role assignment is equivalent to correct syntactic disambiguation, a viable alternative approach is to achieve syntactic disambiguation via semantic analysis. We will now illustrate the limitations of primarily syntactic pathways to syntactic disambiguation based on PP attachment. Let it be said, however, that the architecture subsequently proposed is not limited to the syntactic phenomenon of PP attachment.

2.1 Current NLP Approaches to Structural Disambiguation: PP Attachment

Some of the central techniques for PP attachment disambiguation today include syntactic approaches such as Right Association/Late Closure and Minimal Attachment or syntacto-lexical approaches such as statistical methods and case frames.

With reported attachment accuracies of 92% for German [6] and 94.9% [7] to 95.77% [8] for English, statistical approaches to PP attachment based on machine learning are the most successful and most frequently used ones in large-coverage parsers today. Statistical methods make predictions for PP attachment based on patterns extracted from large text corpora. These approaches favour attachment in agreement with what has been found to exist most frequently in the corpus. Note, however, that the statistical approaches report averages and thus are blind to context, too. Their decision is static rather than contextually dynamic.

A more semantically inclined approach is provided by case frames which constitute an approach to base PP attachment decisions on a formalisation of the verb's syntactic and semantic attachment requirements anchored in the lexicon. With case frames, the attachment decision is based on the verb's syntactic argument selection criteria. These may be extended to include semantic constraints to permit an evaluation of semantic fit for the constituents in question as well. This model is supported by experimental evidence from priming which indicates that syntactic schemas are activated during sentence processing [9]. To our knowledge, however, present-day case-frame implementations do not include extra-sentential context in the assessment of arguments' semantic fit. We therefore highlight the need to extend the case frame model to include cross-modal context as a key component for the assessment of semantic fit in thematic role assignment.

2.2 Challenges in Assigning Thematic Roles

Since a thematic role describes a relation of a noun phrase (NP) to a verb phrase (VP), approaches to thematic role assignment can, in principle, be based on the properties of the NP, the VP or both. As a first step towards a semantically enhanced case-frame approach for thematic role assignment it therefore seems promising to examine the ontological properties of the PP's NP to assess whether the PP can actually fill a specific thematic role, say INSTRUMENT, for a given verb. (S4.1) and (S4.2) clearly illustrate the challenge in this task as the same NP can be a good INSTRUMENT for one sense of the polysemous verb (S4.1) but not for another (S4.2).¹

(S4.1) *He cut the apple off [with a knife.]*_{PP: INSTRUMENT}
Context: Picking an apple.

(S4.2) ?? *He cut her off [with a knife.]*_{PP: INSTRUMENT}
Context: A conversation.

This begs the question whether we can use semantic features of the NP in question to assess its suitability for filling a given thematic role slot. In the literature attempts have been made to define feature-based thematic role hierarchies as a basis for the decision on an NP's suitability in a given thematic role slot [10], [11]. The most prominent – and also the most frequently investigated – NP feature is *animacy* [12] whose importance in human sentence processing has also been shown using brain-imaging techniques such as ERPs [13].

To our knowledge, however, none of the noun-feature-based categorisation approaches to thematic role assignment achieves coverage over a broader range of verbs, let alone generality. We interpret this lack of generalisability of feature requirements on thematic role slots a clear manifestation of the thematic role filler's semantic dependence upon verbal argument structure. In assigning the totality of all thematic selection criteria to the verb, Dowty [14] has disassembled all hopes of constructing a generalised, i. e.: verb-independent, noun-feature-based thematic role hierarchy.

Dowty's view of dominating verb meaning is supported by evidence from priming experiments suggesting that the thematic roles AGENT, PATIENT and INSTRUMENT are intrinsically integrated into the situation schema referenced by the verb [15]. Ferretti et al. found that verb meanings prime nouns which are commonly perceived as good filler nouns for the aforementioned thematic roles. They conclude that thematic role knowledge is closely and inseparably coupled to verb meaning.

In our view this experimental evidence motivates to model thematic role assignment in our architecture with close proximity to verb meaning. We also acknowledge the need to step beyond intrasentential semantic relations in thematic role assignment and to include additional semantic knowledge from utterance context.

¹ Note that we have excluded here the effect upon thematic role assignment arising from the semantic interactions between preposition and the NP in the PP to afford the PP's overall semantic content here.

2.3 The Extent to which Context Influences Thematic Role Assignment

It is not a new claim that context and world knowledge are required for thematic role assignment. McCawley has argued as early as 1968 [16] that selectional restrictions must make reference to world knowledge and that some of this knowledge will be verb-specific. McCawley's *selectional restrictions* effectively are the verb's perspective on thematic role assignment. [2] also argue for a turn away from structural towards a semantic and context-based approach for the resolution of local ambiguity in natural language. Their position arises from experimental evidence demonstrating that thematic role assignment can be context-directed for English reduced relatives [2]. Their findings are further supported by other techniques such as eye-tracking experiments [12].

While the importance of context in thematic role assignment is unchallenged, the question remains just how strong its contribution to ambiguity resolution is. While contextual information can in some cases be the key to fully disambiguating an utterance, in other cases it may lead the hearer astray. Trueswell and Tanenhaus [12] argue to take into account both the strength and the local relevance of contextual constraints. Hence, context information is preference-directing in character. It should therefore not be modelled as a rigid grammatical rule but rather as a preference indicator dynamically adjusting to contextual influences. Our architecture should permit exactly this: the modelling of preference gradients ranging from hard (must be met) to soft (can be but need not be met).

3 Proposed Architecture for Cross-Modal Context Integration

3.1 Conceptual View

Our architectural approach centres around CDG, a weighted-constraint dependency syntax parser downloadable from <http://nats-www.informatik.uni-hamburg.de>. Based on a full-form lexicon and a weighted constraint grammar, CDG produces labelled dependency structures analogous to those in (D1) in a three-step process:

1. The sentence is submitted to a range of external components such as a part-of-speech tagger and a chunk tagger with the request for feedback.
2. When feedback from the external components has been received, CDG applies unary constraints from the grammar. Every constraint violation incurs a penalty. The more important ('the harder') the constraint, the more severe its violation penalty.
3. Once compliance with the unary constraints has been checked, CDG applies the grammar's binary constraints and searches for the dependency structures with the lowest overall penalty score. Penalty scores are the product of all penalty scores incurred during steps 2 and 3. CDG then outputs the best ranked dependency parse.

To improve structural disambiguation for PP attachment CDG currently integrates an external component, the PP Attacher, with the following properties²:

- a. The PP Attacher is integrated in step 1 above. Note that at that stage CDG has not performed any syntactic parsing and hence has no higher syntactic information available yet. The PP Attacher can therefore only process its input at token level.³
- b. Based on attachment observations from large text corpora the PP Attacher assigns scores for the attachment of a preposition to other words in the same sentence.
- c. Scores are provided for attachment to all verbs in the sentence as well for attachment to all nouns left of the preposition.
- d. For large texts the PP Attacher attains an accuracy of 92% [6].

The PP Attacher's properties above have the following implications: a. implies that the current PP Attacher component is ignorant of sentence structure as well as supra-sentential context. b. implies that the scores returned by the predictor are constant for a given preposition-attachment-word pair. This means we can expect exactly one static prediction for the disambiguation of (S1)-like sentences, irrespective of which context they appear in. This is a direct consequence of the fact that attachment probabilities are based on the preposition alone. The PP kernel noun's contribution to the overall PP meaning is presently ignored. c. implies that the PP Attacher evaluates attachment at word- and not at phrase-level. While d. attests unprecedented disambiguation accuracies, properties a., b. and c. indicate that the PP Attacher will not perform as well in dynamically changing contexts.

To enable our architecture to simulate cross-modal context integration as observed in human sentence processing we propose the following enhancements:

1. Enable semantic analysis capabilities in the parser. Specifically, enable the parser to assign thematic roles correctly.
2. Have a plausibility assessment guide thematic role assignment. Perform the plausibility assessment based on cross-modal context information managed in a separate Context Model.
3. Add suitable constraints to the grammar which direct attachment preferences based on the context-sensitively assigned thematic roles. Define these constraints such that parsing accuracy outside of the modelled context domain remains uncompromised.

As for the representation of context information, we believe that the most suitable representation will be in the form of a machine-readable context ontology of cross-modal propositional knowledge. For the modelled domain, it is to contain a represen-

² For a full account of the PP Attacher see [6].

³ Note, however, that the PP Attacher can be combined with other external components capable of making predictions on higher syntactic properties such as phrase structure. The PP Attacher uses the input of those components to have access to higher syntactic information which CDG cannot provide at the early stage in the parsing process when the PP Attacher is called.

tation of all relevant objects and their syntactically significant relationships to each other.

We maintain that the domain-specific representation of extra-linguistic knowledge need not be exhaustive as long as it is sufficient to support structural disambiguation. Further, the nature and origin of the context knowledge are arbitrary. This means that, in principle, the Context Model can function as a unifying interface for a large variety of contextual data. In robotics, for example, the Context Model can be fed with propositional knowledge derived from cross-modally integrated sensor data such as video or audio data. Effectively, this permits the integration of propositional knowledge from low-level sensory data into the syntax-semantic interface.⁴

It is also worth mentioning that we are assuming in this architectural approach that contextual information is indeed congruent with utterance meaning. Strictly speaking, context information is just one cognitively very plausible and frequently utilised guiding principle to direct structural ambiguity resolution. It is, however, not impossible to imagine a scenario in which context and utterance meaning diverge, i. e. a scenario in which context suggests one meaning while the speaker actually intended another meaning for the utterance. Speakers may purposely employ such semantic contrast in order to achieve special communicative effects, e. g. in humorous, ironic or metaphorical speech. The proposed architecture will detect such cases of divergence between utterance and context semantics through violation of the corresponding constraints.

3.2 Implementation View

To realise the conceptual approach presented in the preceding subsection several changes to CDG's existing standard implementation for German need to be made. We now outline these changes in detail.

The existing standard lexicon needs to be extended to include thematic role selection criteria for all verbs within the domain modelling scope. The existing standard grammar needs to be extended to perform thematic role assignment. Thematic roles will be assigned based on a combination of the verb's lexical properties and plausibility scores obtained from a Plausibility Predictor Component (PPC). The PPC manages the communication between the parser and the Reasoning Component (RC) which accesses and queries the Context Model (CM). Queries are formulated by the PPC based on the input sentence material and lexical information it receives from CDG. The context model is created manually using an ontology editor. The aforementioned components integrate into the Context Integration Architecture (CIA) shown in Fig 1.

⁴ For the sake of argument, we here omit the discussion of the additional levels of complexity inferred by the need to perform suitable sensor data fusion prior to integration into the context model. Our assumption is that context information in the Context Model has already been fused across modalities.

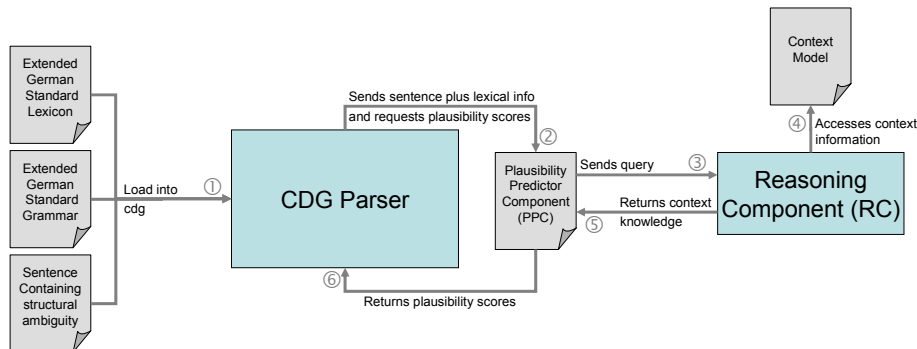


Fig. 1. The Context Integration Architecture (CIA).

Context integration proceeds along the following sequence of steps:

1. The extended German Standard Grammar and Lexicon as well as the input sentence containing structural ambiguity are loaded into CDG.
2. CDG submits the ambiguous sentence plus additional static information from the Lexicon to the PPC and requests plausibility scores for structural disambiguation.
3. Based on the input received from CDG, the PPC formulates a query and submits it to the RC.
4. The RC accesses context information in the Context Model which contains a representation of the sentence context under consideration. The RC also performs reasoning on the context information to produce context knowledge.
5. The RC returns its context knowledge to the PPC.
6. Based on the context knowledge received from the RC the PPC assigns plausibility scores and returns them to CDG. The CDG-internal syntax parsing process now starts. The grammar's integration constraints consider any thematic role assignment preferences arising from the plausibility scores received.

4 Conclusion

In this paper we have outlined a system architecture motivated by context integration effects during human sentence processing. Our architecture couples a reasoning component operating on a context ontology with a semantically enabled syntax parser. Contextual influences are integrated via context-dependent plausibility assessments. Since the Context Model in the architecture is accessed before a sentence is parsed the syntactic structure obtained can respond dynamically to changing context representations. Our architecture therefore permits the simulation of cross-modal context integration in human sentence processing.

We have also hypothesised that integrating cross-modal context into syntactic constraint dependency parsing will significantly and substantially improve the accuracy of structural ambiguity resolution.

5 Future Work

In this paper we have focused on applying the proposed architecture on PP attachment. We will study parsing accuracy for other types of German syntax ambiguity such as subject-object ambiguity or genitive-dative ambiguity in feminine singular nouns.

Another interesting area of study opened by the proposed architecture is constraint relaxation. Here, the effect of systematic modifications to the grammar's constraint penalty scores upon overall parsing accuracy will be investigated.

We also will use the proposed architecture to model more complex phenomena in human sentence processing such as cross-modal compensation. This effect assumes the reliance on contextual information to achieve improved robustness to ungrammatical or incomplete input.

By extending the mechanism for populating the Context Model from manual to automated, processing can extend to continuous data streams, which would permit an expansion from sentence-by-sentence operation to continuous operation. Application domains in which the Context Model is filled with a continuous flow of propositional knowledge obtained from different cross-modal sources (e. g. from a robot's camera and microphone) are of particular interest. By looping contextual representations based on parsing results back into the Context Model the architecture may be extended to build up a discourse history.

References

1. Knoeferle, Pia S. (2005). *The Role of Visual Scenes in Spoken Language Comprehension: Evidence from Eye-Tracking* (PhD Thesis). Saarbrücken: Universität des Saarlandes.
2. Crain, Stephen & Steedman, Mark (1985). On not Being Led up the Garden Path - the Use of Context by the Psychological Syntax Parser. In D. Dowty, L. Karttunen and A. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 320 – 358). Cambridge University Press.
3. Lieberman, Henry, Faaborg, Alexander, Daher, Waseem, & Espinosa, José (2005). How to Wreck a Nice Beach You Sing Calm Incense. *International Conference on Intelligent User-Interfaces, IUI 2005*, January 9 – 12, 2005.
4. Watanabe, Katsumi (2001). *Cross-modal Interactions in Humans* (PhD Thesis). Pasadena, California: California Institute of Technology.
5. Tanenhaus, Michael, Spivey-Knowlton, Michael J., Eberhard, Kathleen M. et al. (1995). Integration of visual and linguistic information in spoken language comprehension. *SCIENCE* (Volume 268), 16 June 1995, pp. 1632 – 1634.
6. Foth, Kilian & Menzel, Wolfgang (2006). The Benefit of Stochastic PP Attachment to a Rule-Based Parser. In *Proceedings of the 21st International Conference on Computational Linguistics*. Sydney: Coling-ACL-2006.
7. Lüdtkke, Dirk & Sato, Satoshi (2003). Fast Base NP Chunking with Decision Trees – Experiments on Different POS Tag Settings. In A. Gelbukh (Ed.) *Computational linguistics and intelligent text processing*, Springer LNCS, 2003, pp. 136 – 147.
8. Kudo, Taku & Matsumoto, Yuji (2000). Use of Support Vector Learning for Chunk Identification. In *Proceedings of CoNLL-2000 and LLL-2000*. Lisbon, Portugal.

9. Auble, Pamela & Franks, Jeffery J. (1983). Sentence comprehension processes. *Journal of Verbal Learning and Verbal Behavior* (22), pp. 395 – 405.
10. Grimshaw, Jane B. (1990). *Argument structure*. Cambridge MA: MIT Press.
11. Simpson, Jane (1991). *Warlpiri morpho-syntax: A lexicalist approach*. Dordrecht: Kluwer Academic Publishers.
12. Trueswell, John C., Tanenhaus, Michael K., & Garnsey, Susan M. (1994). Semantic Influences on Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution. *Journal of Memory and Language* (33), pp. 285 – 318.
13. Kuperberg, Gina R., Kreher, Donna A., Sitnikova, Tatiana, Caplan, David N., & Holcomb, Phillip J. (2006). The Role of Animacy and Thematic Relationships in Processing Active English Sentences: Evidence from Event-Related Potentials. In *Brain and Language*, in press.
14. Dowty, David (1989). On the Semantic Content of the Notion of 'Thematic Role'. In Genaro Chierchia, Barbara H. Partee and Raymond Turner (Eds.), *Properties, types and meaning* (Volume II: Semantic issues). Dordrecht: Kluwer Academic Publishers, pp. 69 – 130.
15. Ferretti, Todd R., McRae, Ken & Hatherell, Andrea (2001). Integrating Verbs, Situation Schemas, and Thematic Role Concepts. *Journal of Memory and Language* (33), pp. 516 – 547.
16. McCawley, J. D. (1968). The Role of Semantics in Grammar. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehart, & Winston.

