Integrating Cross-Modal Context
for PP Attachment Disambiguation

Patrick McCrae

CINACS Graduate Research Group
Department of Informatics, Hamburg University
Vogt-Kölln-Straße 30, 22527 Hamburg, Germany
patrick.mccrae@informatik.uni-hamburg.de

Abstract

Correct prepositional phrase (PP) attachment is a notoriously hard problem in natural language parsing. Despite the inherent structural ambiguity it introduces, human communication still succeeds in most cases with remarkable robustness. The reason for this is that in processing natural language humans also integrate information from sources other than the linguistic material alone. Humans access additional knowledge to enrich the semantic specification that guides syntactic and semantic disambiguation. One important source of additional knowledge for humans is cross-modal context as construed from sensory perception.

Motivated by effects during human sentence processing we propose to integrate contextual knowledge into syntactic parsing to support PP-attachment disambiguation. We hypothesise that integrating cross-modal context into syntactic constraint dependency parsing will significantly and substantially improve the accuracy of PP attachment disambiguation.

Keywords: Natural Language Processing, PP Attachment, Constraint-based Parsing, Context Modelling.

1. Introduction

Considering the complexity of factors contributing to successful natural language interactions, human language processing is surprisingly robust against ambiguous or ungrammatical input – in fact, much more robust than 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 alone. 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 ambiguity of various types. 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 as a resultant of the overall context evaluated.

An extensively studied example for syntactic ambiguity is prepositional phrase (PP) attachment. It is of particular interest because in many cases the resolution of structural ambiguity cannot be achieved on linguistic grounds alone but requires extra-sentential support information. 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. In the second reading, *the 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. One key source of contextual information in human communication is cross-modal sensory perception [4], [5]. Given such cross-modal context information – e. g. by additionally seeing the image of a man looking through the telescope – we have sufficient information to favour one PP attachment over the other and hence arrive at one of the dependency structure representations given in (D1); attachment PP1 is for the INSTRUMENT role and attachment PP2 for the COMITATIVE role, respectively.

(D1)
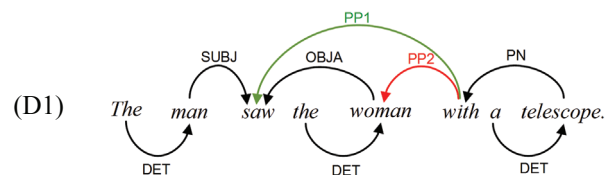
While a large number of NLP techniques for handling PP attachment exist, the majority of these approaches rely on lexical or syntactic properties of the input sentence or statistical distribution patterns of the preposition relative to its attachment constituent. All of these approaches are based on 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 in the syntactic decision process.

Starting out from an existing implementation of CDG, a weighted-constraint dependency syntax parser for German, our goal is to apply our system architecture capable of integrating context to the problem of PP attachment. We hypothesise that integrating cross-modal context information will significantly and substantially improve the accuracy of PP attachment disambiguation.

The subsequent structure of this paper is as follows: We first outline current approaches to PP attachment. We then derive some general requirements for a context-integration architecture supporting PP attachment. In Section 3 we explain how our context-integration architecture meets these requirements and can be applied to the problem of PP attachment, first conceptually, then with an implementation focus.

2. Literature Review

Most heuristics for PP attachment disambiguation in broad-coverage NLP applications today are based on the syntactic or syntacto-lexical properties of the sentence material rather than supra-sentential phenomena such as contextual semantics. We will now illustrate the limitations of some of the key syntactic and lexical approaches to PP attachment disambiguation.

2.1. Current NLP Approaches to PP Attachment

Some of the central techniques for PP attachment disambiguation today include syntactic approaches such as Right Association/Late Closure and Minimal Attachment as well as 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 contextually induced attachment preferences, too. Their decision is static rather than context-dynamic.

A more semantically inclined approach is provided by case frames which base PP attachment decisions on a formalisation of the verb's syntactic and semantic argument requirements anchored in the lexicon. With case frames, the attachment decision is based on the verb's syntactic argument selection criteria. These criteria may be extended to include semantic constraints permitting an evaluation of semantic fit for the constituents. The case-frame concept of verb meaning imposing semantic selection constraints is supported by experimental evidence from priming which indicates that verb-centred syntactic schemas are activated during sentence processing [9]. To our knowledge, however, present-day case-frame implementations do not consider extra-sentential or cross-modal context in their assessment of arguments' semantic fit. The preceding discussion of thematic role assignment (S1) has illustrated just how helpful thematic roles can be for PP attachment. We consequently propose to extend the case-frame model to include context as a key component in the assessment of semantic fit during thematic role assignment for PP attachment disambiguation.

2.2. Challenges in Thematic Role Assignment

Since a thematic role describes a binary relation between a noun phrase (NP) and 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 may therefore seem promising to examine the ontological properties of the PP's NP to assess whether the PP can fill a specific thematic role for a given verb. Attempts have been made in the literature to define feature-based thematic role hierarchies as a basis for the decision on NP-suitability in a given thematic role slot [10], [11]. The most prominent – and also the most frequently investigated – NP-feature is animacy [12] whose significance in human sentence processing has also been shown using ERPs [13].

To our knowledge, however, none of the NP-feature-based categorisation approaches to thematic role assignment has achieved coverage over a broader range of verbs, let alone verb-detached generality. We interpret this lack of generalisability of feature requirements on thematic role slots as a clear manifestation of the role filler's semantic dependence on verbal argument

structure. In assigning the totality of thematic selection criteria to the verb, Dowty [14] has disassembled any further hopes of constructing a generalised, i. e.: verb-independent, NP-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 intrinsically coupled to verb meaning.

For PP attachment this experimental evidence motivates the requirement to model thematic role assignment in close proximity to verb meaning.

2.3. Contextual Influences on Thematic Role Assignment

It is not a new claim that context and world knowledge are required for successful 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. Crain and Steedman [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 PP attachment disambiguation is. Contextual information can be the key to fully disambiguating an utterance in some cases; in other cases again it may lead us astray. Trueswell and Tanenhaus [12] therefore 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 not be modelled as a steadfast rule but rather as a preference indicator that dynamically adjusts to contextual factors. This is of particular importance in PP attachment where in principle both attachments are syntactically (i. e. intra-sententially) possible but extra-sentential factors such as context or world knowledge direct attachment preferences one way or another. To be able to model attachment preferences, our approach to PP attachment must therefore support preference gradients.

3. The Context-Integration Architecture

3.1. Conceptual View

Our contextual approach to PP attachment centres around CDG, a weighted-constraint dependency syntax parser obtainable 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. CDG submits the sentence to a range of external pre-processors (e. g. a part-of-speech tagger and a chunk tagger) with the request for feedback.
2. When feedback from the external pre-processors has been received, CDG checks for satisfaction of the grammar's constraints. This process may include information from the pre-processors. Every constraint violation incurs a penalty. The harder the constraint, the more severe its violation penalty.
3. Once constraints satisfaction has been evaluated, CDG searches for the dependency structure with the lowest overall penalty score. Penalty scores are the product of all penalties incurred in Step 2. 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 properties: [1]

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

Following from a. the 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,

---

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

irrespective of which context they appear in. The PP kernel noun's contribution to the overall PP meaning is presently ignored. From c. it follows that the PP Attacher evaluates attachment at word- and not at phrase-level. The implications of a., b. and c. show that the static nature of the PP Attacher cannot support accurate syntactic decisions when attachment depends on a dynamically changing context. We therefore propose a context-dynamic approach to PP attachment disambiguation by:

1. Assigning thematic roles based on lexical properties.
2. Representing and managing context knowledge in a separate Context Model.
3. Directing thematic role assignment based on contextual plausibility.

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 must 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). The PPC formulates queries based on the sentence material and static information it receives from CDG. We employ RacerPro as a DIG-compliant description-language-reasoner. The communication between the PPC and the RC proceeds across the DIG V1.1 API [17]. The Context Model is created manually using an the Protégé knowledge base editor obtainable from *http://protege. stanford.edu*. The aforementioned components integrate

into the context-integration architecture shown in Fig 1. Context integration into syntax parsing then proceeds along the following sequence of steps:

1. The extended German Standard Grammar and Lexicon as well as the input sentence containing a PP attachment ambiguity are loaded into CDG.
2. CDG submits the sentence tokens plus their thematic role preferences obtained from the Lexicon to the PPC and requests a contextual plausibility score for each of those roles.
3. Based on the input received from CDG, the PPC generates a query and submits it to the RC.
4. The RC queries the CM, a formal representation of the entities in the utterance context as well as their relations to each other.
5. The RC returns its query results to the PPC for scoring.
6. Based on the Context Model properties received from the RC the PPC assigns plausibility scores to every possible thematic role a given token in the sentence can take and returns the scores to CDG. The plausibility scores indicate how likely it will be for the word to take on a thematic role given the context modelled in the CM. CDG then starts the syntactic parsing process. During constraint evaluation CDG assigns thematic roles to the PP and verifies each one of them against the contextual plausibility scores.

4. Conclusion

We have proposed how contextual information can be integrated to support decision making in PP attachment. Contextual influences on attachment decisions can be integrated via contextual plausibility scores. Our approach opens up a contextually aware route to PP attachment via contextually plausible thematic role assignment. We have also hypothesised that integrating context into syntactic constraint dependency parsing will significantly and substantially improve the accuracy of structural ambiguity resolution.
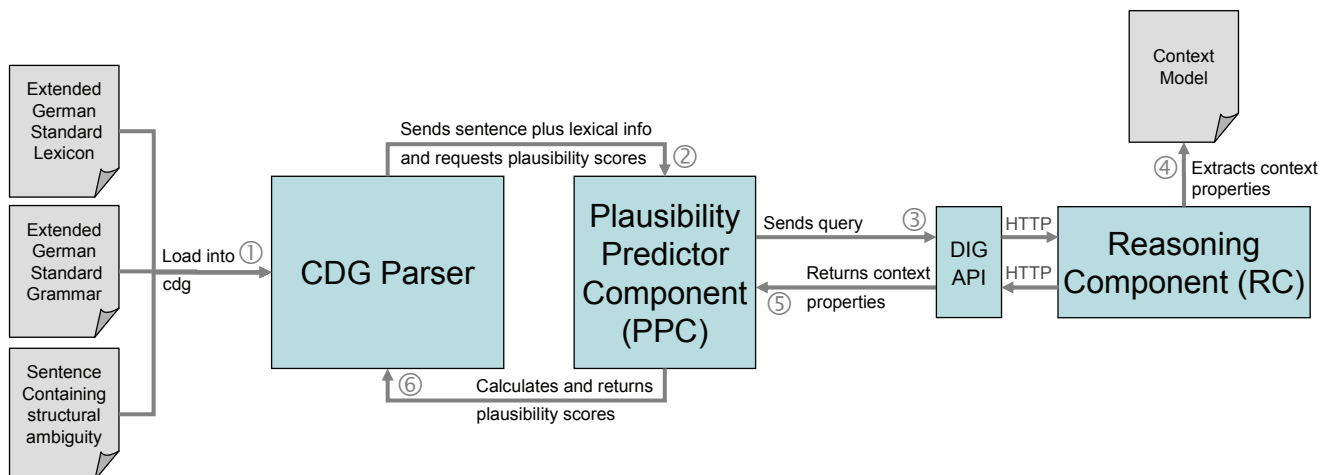

Fig. 1. The proposed context-integration architecture

5. Future Work

Since our context-integration architecture is not specific to the syntactic phenomenon of PP attachment we will also study other types of German syntax ambiguity such as subject-object ambiguity or genitive-dative ambiguity in feminine nouns.

For a quantitative investigation of context upon syntax we will investigate constraint relaxation. Here, the effect of systematic modifications to the grammar's constraint penalty scores upon overall parsing accuracy will be evaluated.

By feeding suitable representations of parse results from preceding sentences back into the Context Model, we plan to consecutively build up a discourse history accessible to the PPC. This discourse history may provide valuable referential information for further improvements in PP attachment. Since in most cases humans, too, have access to a discourse context its integration into the parsing process appears to be a promising step towards more natural parsing scenarios in NLP.

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] Lüdtke, 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), 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 Gennaro 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 a Grammar. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehart, & Winston.

[17] Bechhofer, Sean (2003). The DIG Description Logic Interface: DIG/1.1. In *Proceedings of the 2003 Description Logic Workshop* (DL 2003).