How Reasoning Achieves Context Integration into Syntax Parsing

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— *Motivated by a representational model for the cross-modal interaction between language and other modalities we present a framework for the integration of contextual information into syntactic parsing. We provide a detailed description of the reasoning steps at the interface between a constraint-based parser and a semantic knowledge representation of context information. Our model implements context integration as a three-step process that involves (1) intra-modal grounding, (2) cross-modal mapping and (3) semantic role inferences. We illustrate how these reasoning steps can assist in the task of syntactic disambiguation given biasing context information.*

Keywords:
Reasoning, Natural Language Processing, Syntax Parsing, Context Integration, Cross-Modality

1. Introduction

Cross-modal interactions between language and other modalities have been studied with a primary focus on the interaction between vision and language. Historically, most investigations adopted a behavioural approach and employed the visual-world paradigm, i.e. they studied subjects' response to concurrently presented visual and linguistic stimuli. Despite the profound insights that have resulted from these efforts over the last three decades or so, modelling approaches for the cross-modal interaction of language with other modalities in articifial systems are still rare and few.

We present a framework implementation based on the context integration archtitecture proposed by McCrae and Menzel [1] which employs a sequence of reasoning steps to integrate context information into the process of syntactic parsing. We motivate in detail the different types of reasoning steps required for context integration into language processing and describe how those steps have been implemented in our framework. The key benefit of applying the framework lies in its power to resolve syntactic ambiguities based on contextual extra-sentential information.

We have structured this paper as follows: In Section 2 we provide a brief overview over the cognitive motivation for our framework. In Section 3 we outline the overall process of context integration in our implementation. Section 4 discusses the individual reasoning steps in detail. We conclude with a summary of our main points and an outlook for future work in Section 6.

2. Cognitive Motivation

While humans automatically and unconsciously integrate the totality of information sources available at the time of processing a (potentially ambiguous) utterance, be it world knowledge, discourse context, visual scene context or others, the implementation of intelligence to perform analogous integration processes in an artificial system remains a non-trivial task. McCrae [2] describes the integration of a semantic knowledge representation of visual scene context into syntactic processing. There is no reason to assume, however, that the generation of contextual knowledge representations should be limited to cross-modal context. The representation is based on concept instances connected by semantic relations. It hence is not modality specific and could be used to encode information from other kinds of contexts as well. The framework mechanisms we are about to outline are capable of processing any kind of extra-linguistic information as long as it is semantically encoded in the representational format presented. In the following, we refer to arbitrary external information that can be encoded in terms of concept instances related by semantic relations as 'context'.

Based on a wide range of linguistic and cognitive examples Jackendoff argues for an interaction between language and other modalities at a conceptual, semantic level of representation [3], [4]. Jackendoff introduces *Conceptual Structure* as the mental level of representation at which language interfaces with all other, non-linguistic modalities – be they sensory or representational in nature. Interactions occur between concepts, concept instances and semantic relations, all of which are accessible to reasoning. Our framework models the interaction of context with language as mediated by Conceptual Structure by including implementing:

- an ontology of concepts (T-box)
- a knowledge representation of context containing a set of concept instantiations linked via thematic relations (A-box)
- a reasoner (FaCT++, obtainable from [5])
- an integrated semantic representation built up by the

parser based on input from (1) the current state of its syntactic representation of the linguistic input and (2) the semantic relations found in the contextual knowledge representation

3. The Process of Context Integration

McCrae and Menzel [1] describe an architecture for the integration of cross-modal context that centres around WCDG, a weighted-constraint parser for German [6]. WCDG provides a generic interface for integrating external information into the parse process by providing scores for dependencies between words prior to the commencement of the actual parse process. We use this mechanism to provide external scores for semantic relations. Scores are calculated based on the compatibility of a given semantic relation with a semantic knowledge representation of contex. When the parser scores a thematic dependency integration constraints provide the predictor scores for that dependency's contextual compatibility. When the dependency is included into the parser's contextually integrated semantic representation, the dependency score contributes multiplicatively to the integrated representation's total score. This mechanism passes dependencies with a score of '1' and excludes those with a score of '0'. Graded preferences are expressed by assigning a fractional score between 0 and 1.

In our implementation of McCrae and Menzel's architecture all words in the input sentence are handed over to the context integration predictor. To obtain contextual scores for the semantic dependencies between a pair of words we need to pass through a sequence of three steps which we will outline in detail in the following sections.

1) Intra-Modal Grounding: Mapping each word onto its set of activated concepts.
2) Cross-Modal Mapping: Determining which words in the linguistic input map onto which concept instances in the contextual knowledge representation.
3) Semantic Reasoning: Scoring semantic dependencies for other dependant-regent pairs based on the contextually asserted semantic relation.

4. Reasoning Steps

4.1 Intra-Modal Grounding

Intra-modal grounding is a procedure which, by definition, is only required for sensory modalities in which sensory input needs to be associated with a corresponding concept in Conceptual Structure. Representational modalities, in contrast, are already represented symbolically and therefore have been grounded in the process of creating their symbolic representation.

In our model, the mapping between a word and its associated, or activated, concept is mediated by the word's lexical base form which is one of the elementary entries in each surface string's lexicon entry in WCDG's full-form lexicon. Nouns are mapped via their nominative singular form to their corresponding concept in the ontology. Verbs are mapped via their infinitive onto a situation concept whose lexicalisation is identical to the verb's infinitive. Furthermore, the situation concept's arity and the verb's semantic valence need to coincide, e.g.: The string sehen/VVINF *see* lexicalises two distinct situation concepts in the ontology: SEHEN-ag and SEHEN-ag.th. The former represents a *seeing*-situation in which only an AGENT participates while the latter represents a *seeing*-situation in which both an AGENT and a THEME are involved. Since WCDG's lexicon provides unique lexical details for every surface string, we know the semantic valence of each verb form and can hence establish the mapping to the situation concept with the appropriate situation arity.

Thus far, our argument may be read to suggest that each verb maps onto a single concept. For a robust handling of phenomena such as lexical ambiguity or homophony it is helpful, however, if a single surface string can activate an entire set of concepts. Every surface string in the input sentence therefore maps to a set of concepts whose lexicalisation is equal to the string's lexical base form (cf. Figure 1).

Mannes ⟼ *Mann* ⟼ {MAN}
GenSg NomSg

sah ⟼ {SEE-ag, SEE-ag.th}

Since:

sah ⟼ *sehen* ⟼ {SEE-ag}
VVFIN-Past-ag VVINF-ag

sah ⟼ *sehen* ⟼ {SEE-ag.th}
VVFIN-Past-ag.th VVINF-ag.th

Fig. 1: Mapping surface strings in linguistic input to sets of activated concepts in Conceptual Structure.

The contextual knowledge representation contains concept instances linked by thematic relations. Conceptual grounding has already been performed explicitly by asserting the concept instances in the knowledge representation. At cognitive level, we are assuming that the contextual information has been processed beyond its sensory level and has been interpreted and projected as a semantic representation in Conceptual Structure.

4.2 Cross-Modal Mapping

The aim of context integration is to utilise thematic relations defined in the contextual knowledge representation to assist in dependency assignment for the linguistic input. We are thus using extra-linguistic information to assist our linguistic decision making. One of the key assumptions underlying our model is that a thematic relation θ connecting concept instance I_1 to concept instance I_2 in the context representation can affect the thematic dependency decision between a dependant word w_1 and its regent word w_2 in the parser if and only if the set of concepts activated by w_1 contains concepts compatible with the concept instantiated by I_1 and, analogously, the set of concepts activated by w_2 contains concepts compatible with the concept instantiated by I_2. We hence map words to contextual concept instances based on the compatibility of the concepts activated by a word with the concepts instantiated in context.

As illustrated in Figure 3 this mapping need not be one-to-one. To achieve robust processing of referential ambiguities in cross-modal mapping we adopt a set-based approach. A word in the linguistic input hence maps to a set of concept instances and can engage without penalty in all semantic relations that have been asserted in the context representation for its associated concept instances.

Since the predictor operates in the environment of a constraint-satisfaction formalism it can only influence the parser's dependency assignments via a veto on edges that are inconsistent with the asserted contextual information. The more concept instances a word maps to, the more semantic relations it can engage in and the less restrictive the contextual information is going be upon the resulting integrated semantic representation in the parser. Our experimental findings show that even in the presence of several words in the linguistic input mapping onto multiple concept instances in the context representation the desired disambiguating effect of context upon syntax is achieved. In Section 5 we discuss an example in more detail.

The example in Figure 3 illustrates a special case in which we can perform a further step of inference. Concept compatibility based on gender information tells us that the string *Moderator* can only map to the concept instance HUMAN_01. Based on a procedure of exclusion, we can remove the concept instance HUMAN_01 from the mapping set of the string *Filmstar* and thus obtain an unambiguous one-to-one mapping of words to concept instances. Effectively, this step reduces the number of concept instances that the word *Moderator* maps to. This elimination of word-to-concept instance mappings results in more strongly constraining contextual information.

The assignment of a verb-centred thematic role such as AGENT, THEME or RECIPIENT [7] always involves a compatibility check of the specific situation concept activated by the verb's surface string against the situation represented in context. To preempt a discussions on modelling granularity of situation verb concepts in the context representation we choose to underspecify the situation concepts in our context models. Figure 3 shows that no specific situation concept has been asserted. Rather, a generic ternary event has been instantiated of which nothing more is known than that it involves an AGENT, a THEME and a RECIPIENT. While the framework does permit to model specific situation concepts, we found that the influence of context upon linguistic processing can be achieved with these generic concept instances more robustly. The determining factor in the context of this reasoning step is not primarily the specific ontological class of the situation concept instantiated in context merely its situation arity, i.e. the number of thematic participants that it binds.

If we take the context model to represent a visual scene context, this generalisation has the charme that we do not need to extract specifically, what kind of situation concept is being instantiated in visual context (it would, for example, be difficult, if not impossible, to differentiate between a situation instantiation of SHOW-ag.th versus DEMONSTRATE-ag.th). In our model it suffices to specify that the observed visual context provides an instantiation of some BINARY.SITUATION. Clearly, we can be more specific – but do not need to be.

4.3 Semantic Role Inferences

Now that the full mapping between words in linguistic input and the concept instances asserted in context has been achieved the predictor can utilise this contextual information to compute scores for a range of thematic dependencies between a dependant w_1 and its regent w_2. The computation of scores is a multi-step process that we outline in detail in the following subsections.

4.3.1 Admit Contextually Asserted Thematic Relations

Suppose that a dependant w_1 and its regent w_2 map to a pair of concept instances between which an AGENT relation has been asserted in the context representation. Based on the positive evidence of an AGENT relation in context the predictor will assign a score of 1.0 to the AGENT dependency from w_1 to w_2, which corresponds to full permissibility of that dependency. Note that this is not to say that the dependency will appear in the final parse structure proposed as a solution by the parser. It only says that if it appears, it does not incur a penalty score. Admitting a dependency as such is not too strong a statement yet since the predictor will also deliver a score of 1.0 for any dependency edge for which it cannot make a concrete prediction due to limited or unavailable information in the context representation. The latter scores are implicit permissions since they are not based on concrete positive evidence in the context model.

With the admission of the contextually detected thematic relation, however, we can rest assured that between w_1 and w_2 the AGENT dependency will not be suppressed by the parser based on context information. From the point of view of reasoning, the important steps follow now because we can infer a range of vetoes on other thematic dependencies,

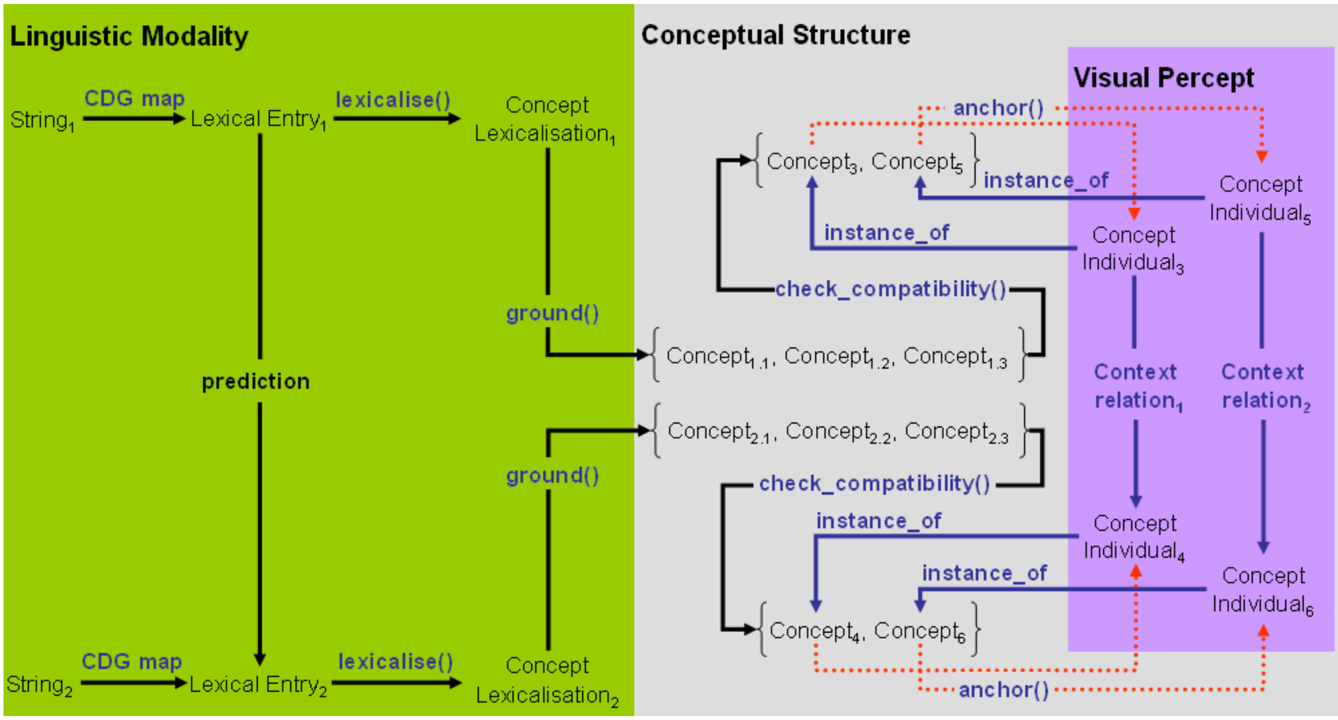

Fig. 2: Mapping procedure in cross-modal integration.

Language
Der Moderator überreicht dem Filmstar einen Oscar.
The male mc hands over an Oscar to the movie star.

Context Model

$$\text{HUMAN\_01} \xrightarrow{\text{is AGENT for}} 3^{ary}.\text{SITUATION\_01}$$
$$\wedge\ \text{WOMAN\_01} \xrightarrow{\text{is RECIPIENT for}} 3^{ary}.\text{SITUATION\_01}$$
$$\wedge\ \text{PRIZE\_01} \xrightarrow{\text{is THEME for}} 3^{ary}.\text{SITUATION\_01}$$

Cross-Modal Mapping

$$\textit{Moderator} \xrightarrow{\text{compatible with}} \{\text{HUMAN\_01}\}$$
$$\wedge\ \textit{Filmstar} \xrightarrow{\text{compatible with}} \{\text{HUMAN\_01}, \text{WOMAN\_01}\}$$
$$\wedge\ \textit{Oscar} \xrightarrow{\text{compatible with}} \{\text{PRIZE\_01}\}$$

Fig. 3: Cross-modal mapping of surface strings in linguistic input to sets of contextual concept instances in Conceptual Structure.

which makes the predictor component a powerful source of constraining context information for linguistic processing in the parser. If the penalties delivered by the predictor are hard penalties, i.e. penalties which inflict a 0 score, they will prevent the entire tree from being included in the searched solution space.

4.3.2 Veto based on the Uniqueness of Thematic Roles

The second important guideline for the assignment of thematic dependencies in the parser is the premise that thematic roles within a situation are unique and mutually exclusive. Based on this rule, our framework assigns penalties to all non-AGENT dependencies between w_1 and w_2. These additional vetoes can subsequently be overwritten by explicit permissions in case the predictor detects positive evidence for another thematic relation between w_1 and w_2 in the context representation. Observe that this procedure also has implications on how we model reflexive situations in our representation of contextual information. For a situation like *Bennet is buying himself a present* we resolve all references and mark co-reference by co-indexation as shown for the assignment of AGENT and THEME relations in Figure 4.

Context Model

$$\text{BENNET\_01} \xrightarrow{\text{is AGENT for}} \text{BUY-ag.re.th\_01}$$
$$\wedge\ \text{PRESENT\_01} \xrightarrow{\text{is THEME for}} \text{BUY-ag.re.th\_01}$$
$$\wedge\ \text{BENNET\_01} \xrightarrow{\text{is RECIPIENT for}} \text{BUY-ag.re.th\_01}$$

Fig. 4: Context modelling for reflexive situations.

4.3.3 Veto based on a Single Dependency per Word and Level

The third set of inferences we draw originates from the formal requirement in WCDG's dependency formalism which demands that every word only have one regent on a given level of analysis, i.e.: a word can only enter a single dependency per level of analysis. To satisfy this requirement, the predictor places a veto on all other thematic dependencies originating from the dependant w_1 for which the detected contextual thematic dependency has been permitted. All of these vetoes can be overwritten based on positive evidence subsequently encountered in the context representation.

4.3.4 Veto based on the Closed-World Assumption

The fourth inference we draw from the detection of an asserted thematic relation between w_1 and w_2 originates from the assumption that the modelled interaction between linguistic processing and context occurs in a closed world. We assume that only those contextual entities and relations which have projected into conceptual structure – and hence have been represented in the context model – can and will have an effect upon the interaction between context and language. This is consistent with our initial assumption that the context model represents constituents of conceptual structure. Clearly, only those percepts which actually have been projected into conceptual structure can subsequently be processed further and may result in an interaction with syntax.

Additional information that has not been modelled in the context model but becomes available at a later point in time t_2, say, clearly cannot affect the interaction at a time t_1 which is assumed to precede t_2 temporally. In our model we adopt the view that things have not cognitively registered (yet) do not affect the way context interacts with the processing of language. If, for example, the number of participants detected in a co-present visual scene changes over time, this results in a modification to the representation of visual context – and different representations may produce different interactions between vision and language. It is very plausible that, at a given point in time, only those aspects contribute to the interaction which have a projection into Conceptual Structure – and thus have been asserted in the contextual representation.

As for scoring, the closed-world assumption permits us to impose vetoes on all other thematic dependencies in the sentence that originate from the same dependant and point to any other regent. The veto for another dependency can subsequently be overwritten again in case positive evidence for that dependency is detected in the context representation.

5. Results

While a detailed discussion of the application of the framework to these classes of syntactic ambiguities is beyond the scope of this paper, we will discuss how the framework processes a globally ambiguous sentence representative of one particular ambiguity class. The sentence selected is from a study of Genitive-Dative ambiguity in German feminine nouns [8]. To improve comparability of the sentences we have normalised their introductory main clauses to 'Er weiß, dass ...' *He knows that ...* in all sentences. Consider Sentence (1) (ambiguous constituents highlighted).

(1) Er weiß, dass die Ärztin ***der Patientin*** den Leidenden präsentierte.

He knows that ...

a. Binary Situation (Genitive reading)
... the female patient's female doctor presented the male sufferer.

b. Ternary Situation (Dative reading)
... the female doctor presented the female patient the male sufferer.

The parser's default analysis of Sentence (1) in the absence of contextual information is the Dative reading which corresponds to the syntactic structure shown in Figure 6. We can, however, modulate the parser's dependency assignments by integrating a context representation shown in Figure 5 which corresponds to the Genitive reading in (1) b.

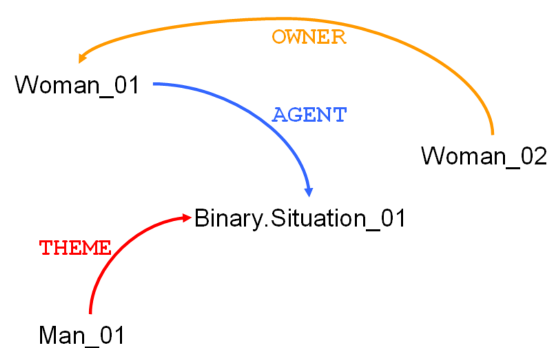

Fig. 5: Representation of visual context corresponding to the binary situation (Genitive reading) of sentence (1).

The result of integrating the binary[1] visual context into the parsing of (1) is the syntactic dependency structure in

[1]We refer to a context as *n-ary* if its central situation concept engages in n direct thematic relations with contextual participants. Note that a context model may, of course, contain more than n entities, not all of which entertain a direct thematic relation with the situation instance.

Figure 7. Integration of a binary context has succeeded in overriding the parser's default ternary situation analysis which previously was obtained in the absence of a context.

6. Conclusions and Outlook

In this paper we have outlined the reasoning procedures at the heart of our framework for the integration of contextual information into syntactic parsing that centres around the weighted constraint parser WCDG. Motivated by Jackendoff's theory of conceptual semantics for the interaction between cognition and language our framework uses contextual information to modulate syntactic decisions. We have argued for a process of context integration consisting of three major steps, namely intra-modal grounding, cross-modal mapping and semantic role inferences. Our semantic role inferences were based on the uniqueness of thematic roles in a situation, the single-regent requirement of WCDG's dependency formalism and the closed-world assumption.

Our future work will concentrate on the effect of constraint relaxation upon context integration and the systematic study of the effect of perceptual uncertainty onto context integration. A more conceptual area of research will be the investigation of how cross-modal mapping preferences based on degrees of conceptual similarity can be included into the present model. This effort will comprise the implementation of a suitable measure for conceptual similarity in our ontology.

References

[1] P. McCrae and W. Menzel, "Towards a system architecture for integrating cross-modal context in syntactic disambiguation," in *Proceedings of the 4th International Workshop on Natural Language Processing and Cognitive Science, Madeira (NLPCS 2007)*, B. Sharp and M. Zock, Eds. INSTICC Press, 2007, pp. 228–237.

[2] P. McCrae, "Integrating cross-modal context for PP attachment disambiguation," in *Proceedings of the 3rd International Conference on Natural Computation, Haikou (ICNC 2007)*, vol. 3. Los Alamitos, CA: IEEE, 2007, pp. 292–296. [Online]. Available: portal.acm.org/citation.cfm?id=981828

[3] R. S. Jackendoff, *Semantics and Cognition*. Massachusetts Institute of Technology, 1983.

[4] R. Jackendoff, *Semantic Structures*. MIT Press, Cambridge, Massachusetts, 1990.

[5] FaCT-PlusPlus Download Page, http://code.google.com/p/factplusplus/, 2009, link verified: 20 April 2009.

[6] CDG Download Page, http://nats-www.informatik.uni-hamburg.de/view/CDG/DownloadPage, 2009, link verified: 20 April 2009.

[7] T. R. Ferretti, K. McRae, and A. Hatherell, "Integrating verbs, situation schemas, and thematic role concepts." *Journal of Memory and Language*, vol. 44, pp. 516–547, 2001.

[8] A. van Kampen, "Syntaktische und semantische Verarbeitungsprozesse bei der Analyse strukturell mehrdeutiger Verbfinalsätze im Deutschen: Eine empirische Untersuchung," Ph.D. dissertation, Freie Universität Berlin, 2001.

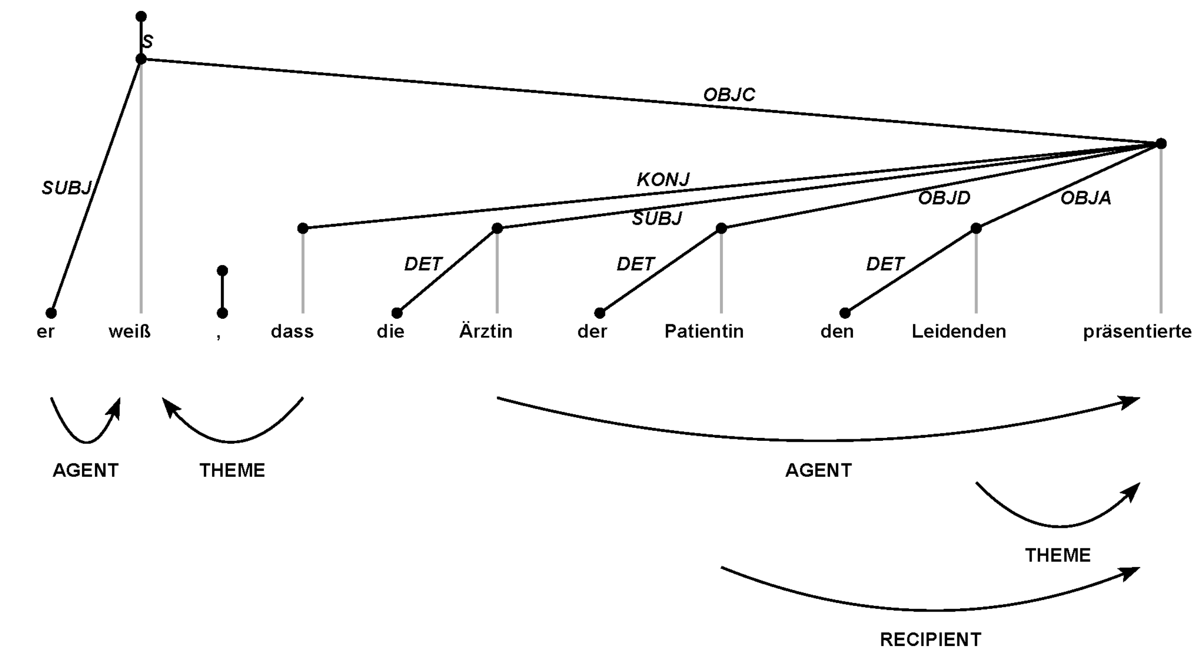

Fig. 6: The parser's default representation in the absence of visual context (Dative reading).

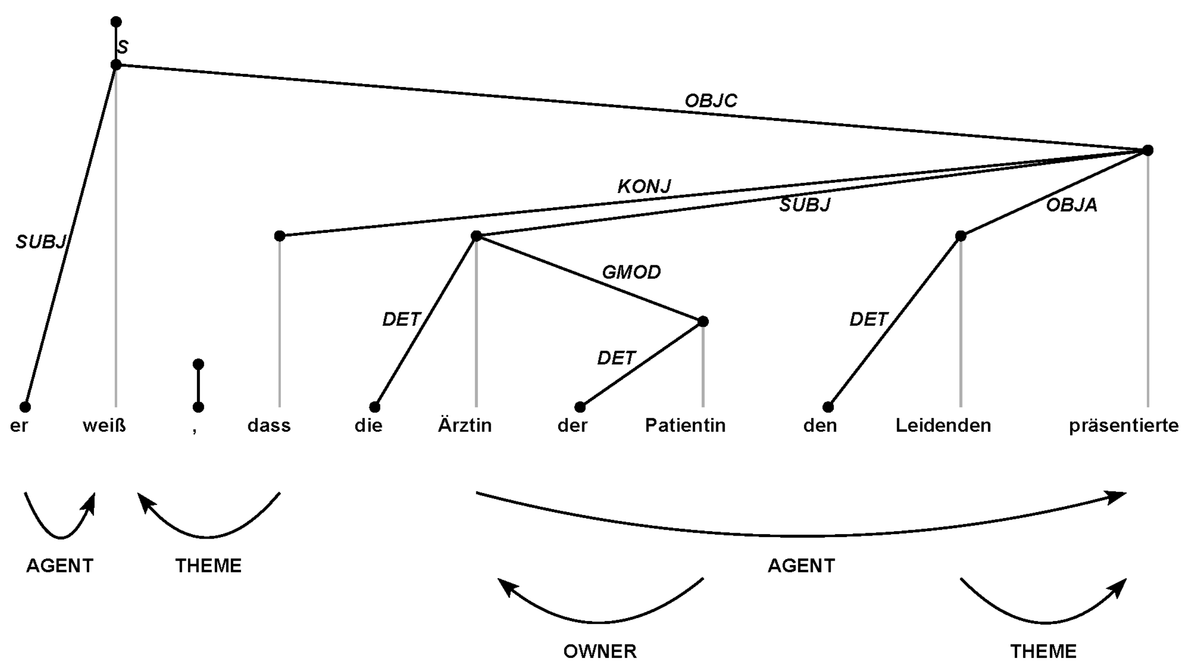

Fig. 7: The parser's cross-modally integrated representation with binary visual context.