EU FP7 CogX
ICT-215181
May 1 2008 (52months)

DR 6.3:
Adaptive extendable grammatical processing

Geert-Jan M. Kruijff, Fai Greeve, Miroslav Janíček, Bernd Kiefer, Hans-Ulrich Krieger

DFKI GmbH, Saarbrücken
⟨`gj@dfki.de`⟩

Due date of deliverable:	July 31 2011
Actual submission date:	July 29 2011
Lead partner:	DFKI
Revision:	final
Dissemination level:	PU

In previous years, WP6 investigated how situated dialogue could be used in human-robot interaction to help the robot learn more about its environment. This involved grounding dialogue in multi-agent models of beliefs and intentions, dealing with the uncertainty and incompleteness in these models, and communicating about the content in these models at different levels of granularity. For these tasks, the robot was always provided with sufficient a priori dialogue competence to carry them out. In Year 3, WP6 explored how such competences could be acquired, as a form of self-extension (Task 6.5). In DR6.3 we describe computational approaches to language learning, acquiring communicative competences from the ground up. This work is closely related to that described in DR6.4. There we explore approaches for extending existing linguistic competence, to verbalize newly acquired conceptual structures (Task 6.6).

Executive Summary

One of the objectives of CogX is self-extension. This requires the robot to be able to actively gather information it can use to learn about the world. One of the sources of such information is dialogue. For this to work, the robot needs to be able to establish with a human some form of mutually agreed-upon understanding, a *common ground*. The overall goal of WP6 is to develop adaptive mechanisms for situated dialogue processing, to enable a robot to establish such common ground in situated dialogue.

In Year 1, WP6 investigated how a robot could carry out a situated dialogue with a human, about items in the world it needed to learn more about. The robot was able to formulate questions against a multi-agent model of situated beliefs, indicating what it did and did not know – and what it would like to know. The robot was able to represent and reason with uncertainty in experience, but it was relatively fixed in the strategies it would follow to communicate with the human about resolving the uncertainty.

The dynamic, interactive setting of CogX in which a robot actively learns requires more than following a fixed, "universal" policy. Learning more, dynamic situations, and the changes in common ground this implies, all require the robot to *adapt* how it acts and interacts, if it is to successfully communicate with a human over time. In Year 2, WP6 investigated several issues in how to make dialogue behavior more *adaptive*. This covered several aspects: (1) Making dialogue strategies more adaptive, and (2) varying how much a robot needs to describe to be optimally transparent.

Throughout Years 1 and 2 we assumed the robot to have a fixed set of communicative competences, particularly where it concerned grammatical resources. Practically this meant that, even though the robot was still learning more and more about the world, it already knew how to talk about it. In Year 3, we let go of this assumption, taking the CogX objective of self-extension to the realm of situated dialogue processing as well. WP6 addressed self-extension for situated dialogue in two aspects. The first aspect concerns the verbalization of large ontologies for modeling common sense indoor knowledge. The categories in these ontologies cover both logical and probabilistic information. Self-extension here focuses on how these categories can be associated with another ontology, namely an ontology of words: WordNet. This provides the means to acquire mappings for verbalizing acquired concepts at varying levels of abstraction and specificity – and in general, it provides a powerful means for large-coverage resources for communicating about indoor environments (Task 6.6). DR.6.4 describes this work in more detail. Deliverable DR.6.3 takes a further step back, broadening the scope of self-extension in situated dialogue to the aspect of language acquisition per se. Looking at the acquisition of language as always taking place against a situated, social, and personal background, WP6 explored how language acquisition for robots could be phrased in a way that takes all

these dimensions of language form to language function and use into account – rather than just "syntactic structure." DR.6.3. presents two semiotically oriented computational approaches to language learning. These approaches are strongly inspired by how children acquire language. They make it possible for the robot to acquire linguistic knowledge that ties words and the structure of expressions together with their situated meaning and socially interactive use, in an integrated fashion (Task 6.5).

Role of situated dialogue in CogX

CogX investigates cognitive systems that self-understand and self-extend. In some of the scenarios explored within CogX such self-extension is done in a mixed-initiative, interactive fashion (e.g. the George and Dora scenarios). The robot interacts with a human, to learn more about the environment. WP6 contributes situated dialogue-based mechanisms to facilitate such interactive learning. Furthermore, WP6 explores several issues around the problems of self-understanding and self-extension in the context of dialogue processing. Dialogue comprehension and production is ultimately based in a situated, multi-agent model the robot builds up. This model captures epistemic objects like beliefs, intentions and events, in a multi-agent fashion. Such epistemic objects cover both situated and cognitive aspects, and already at this level we see forms of self-understanding and self-extension. In Year 3 we take the notions of self-extension and self-understanding into the domain of situated dialogue itself. We show how the generally prevalent view in CogX on cognition, namely its cross-modal character in understanding the world and deciding how to act in there, naturally extends to the way we can perceive of acquiring competence in situated dialogue processing.

Contribution to the CogX scenarios and prototypes

The computational approaches for language acquisition reported in DR.6.3 do not immediately contribute to the CogX scenarios and prototypes – they are still at the level of theory formation. They do however help achieving the overall CogX objectives, in formulating theories about self-understanding and self-extension. Furthermore, one of the approaches is explicitly based in the situated multi-agent models of beliefs and intentions WP6 has been exploring, and which have been integrated in the CogX Dora and George systems.

The approaches for verbalizing conceptual structures at varying levels of abstraction and specificity reported in DR.6.4. are integrated into the CogX Dora and George systems. They are employed to help verbalize the robot's understanding of the environment, and to communicate surprises and their underlying reasons.

1 Tasks, objectives, results

1.1 Planned work

The overall goal of CogX is to arrive at a theory of cognitive robots which are capable of self-understanding and self-extension. During the last years, WP6 worked on adaptive mechanisms for situated dialogue processing that would enable a robot to discuss with a human what it did and did not understand about the world. And, thus, through such dialogue, it could gain more information to help it learn more.

These efforts always started from the assumption that linguistic resources necessary to talk about the world were available to the robot. The issues WP6 addressed in Years 1 and 2 concerned not so much where these resources were coming from, but rather how they could be used – and how that use could be adapted to optimally fit the context and intentions of the agents involved. Some of the decision processes involved in this were acquired off-line (POMDP models for adaptive dialogue management), and an initial model for controller-based adaptive dialogue processing was proposed.

The planned work for WP6 in Year 3 is to take the issue of self-extension further, into the domain of situated dialogue processing. In the current deliverable, DR.6.3, we address Task 6.5:

Task 6.5 *The goal is to investigate how existing lexico-grammatical knowledge can be extended to cover novel categorical knowledge. We will develop methods for learning two types of mappings: a mapping relating a word's lexical meaning to a predicate-argument structure based on the associations of the category this meaning reflects, and a mapping relating a word's predicate-argument structure to a syntactic family that can express the structure.*

The intention behind Task 6.5 was to see how an existing grammar could be extended with new words, and mappings between those words and meaning representations such that they could be grounded in the knowledge representations of the robot. We have taken that intention two steps further: One, as we report in DR.6.4, we have lifted the task from individual words and categories to ontologies of (common sense) knowledge, and of words and their categorical organization. Second, we have taken a step further by taking a step back, and truly address the issue of self-extension – by starting to address the problem of how a robot could acquire "linguistic knowledge." How we have done that, and what we precisely understand by "linguistic knowledge" we explain in §1.2 below.

1.2 Actual work performed

1.2.1 Language learning

In CogX, we take a cross-modal perspective on cognition. Multiple modalities interoperate to aid in processing experience, acquiring new knowledge, deciding how to act. We have followed out the same view when it comes to situated dialogue processing, as we already argued for in [39]. Understanding and producing situated dialogue begins and ends with context: The situated, and socially interactive context within which the dialogue takes place, and the system's "personal" short- and long-term experience into which it grounds this dialogue. For Task 6.5 we have taken this perspective to the issue of self-extension in situated dialogue itself – i.e., language learning.[1]

Typically, computational models focus only on a very specific aspect of learning language, like syntactic structure ("grammar induction") or verbal argument structure ("verb frame/class induction"). Furthermore, learning is mostly supervised, working with corpora annotated with information about the structure and/or meaning to be learnt for expressions.

None of that fits with the CogX perspective, nor with the settings within which we would like robots to learn language. Robots interact with humans and with an environment they only partially understand (incompleteness, uncertainty). They learn as they go. Self-understanding and self-extension imply that even when a human may name an object, or object properties, or an event, the robot may not immediately be able to resolve that name to something it experiences (but may not understand in and by itself). So whatever supervision there may be, it is at most a guidance, and even so subject to (referential) uncertainty. How a robot is to structure its experience and understanding internally is unsupervised, (at least in the sense of machine learning). Yet, having said that, it is also immediately clear that a robot is not learning aspects of language in isolation. It does not first do syntax, then semantics, then pragmatic use. Such a horizontally stratified modularization makes very little sense as it would keep the robot from being actively involved in the interaction.[2] Before it could ask a question, it would first need to understand all there is to understand about meaning. And before that, before it could contemplate meaning, it would first have to know all there is to know about syntax.

Unfortunately, this type of modularization still appears to be fundamental to most linguistic theory. And it is strangely at odds with what we currently believe we understand about how children acquire language. As children develop biologically, cognitively, and socio-emotionally, there is a

---

[1] Hereafter we follow Seginer [57] in using the term *language learning* for models of how a computational system could acquire language, and *language acquisition* to indicate the real-life human processes.

[2] See also the arguments Brooks advanced against horizontal modularization in cognitive systems advanced in the 1980's and early 1990's, [11].

close interaction between these developments, and their individual development of language – in its form, the meaning these forms can express, and how to use these forms and meanings communicatively to specific ends. And children go through this development while interacting with the world, and with other children and adults. Children learn, individually but also from others, through others, with others.

This raises many fundamental questions about the nature of meaning, of the construction and use of meaning, of its acquisition. There is no denying in that there are different aspects to meaning. But both in psycholinguistics and neurolinguistics we find a plethora of evidence of integrative effects that we find across these aspects, when people process (spoken) language (cf. [39]). Language processing is inherently cross-modal, and for all we can say and see at the moment, this naturally already starts when children acquire language [74, 25].

Therefore, rather than following out "traditional" unsupervised computational approaches to language learning, we have explored alternative ways of formulating and addressing the problem of language learning that is closely in line with observations on child language acquisition: Namely, that language learning is situated, socially interactive, and personal.

Language learning in robots as a situated, socially interactive, personal process. *A robot cannot learn language simply by learning words. It needs to be able to acquire language within interactions with the world, and with other agents – hence, with words come situated meanings, ways in which words and meanings can be composed into larger complexes, and how these meanings are used communicatively to build up common ground. Kruijff (Annex 2.1) and Greeve (Annex 2.2) explore sign-based approaches to model aspects of language learning in ways that are closely inspired by insights in child language acquisition. Greeve (Annex 2.2) adopts Language Games (cf. Steels) to explore how semantic constructions for spatial configurations can be acquired by an inexperienced agent, through situated interaction with a more knowledgeable agent. This framework is based on semiotic networks for connecting language and categorical knowledge, and focuses specifically on the use of Fluid Construction Grammar in learning language through Language Games. Kruijff (Annex 2.1) formulates a novel approach in which signs relate words or expressions immediately to their situated grounding, mediated by their effect on a multi-agent model of beliefs and intentions. The approach applies a cross-modal, process-based perspective on the acquisition and use of (structured) sign processes in situated, social dialogue between a learning robot, and a human. It combines insights in large-coverage grammar induction, with socially interactive learning in human-robot interaction.*

The approaches reported of here are both based on the view that child language acquisition is a cross-modal process in which the entirety of social context, linguistic structure, and (personal) categorical understanding of the world come together to make sense of communication, and to guide its further development [7, 8, 74, 25]. This leads to a type of model which is ecological, taking the dynamics of both the individual and its environment into account in formulating a learning process [42, 20], and an understanding of meaning which is essentially embodied, in some sense metaphorical for experience [40, 41]. The more structural characterization of these processes can be based on semiotic sign structures, dating back to American Pragmatism (Peirce, Mead) and Prague School functionalist linguistics (Jakobson). See also §1.3 for more connections to the state-of-the-art.

1.2.2 Additional work performed

The main work reported of in this deliverable concerns the efforts in language learning, within WP6. Besides language learning, we have also worked on other (related) aspects. Particularly Kruijff (Annex 2.1) applies the intentional, continual perspective on situated dialogue processing which has been developed in WP6. Here, we have extended our approach to deal explicitly with uncertainty *and* incompleteness, i.e. knowledge gaps.

Partiality in abductive inference for dialogue processing *Both uncertainty and incompleteness are pervasive throughout all levels of understanding a robot builds up – including understanding situated dialogue. Earlier work in WP6 already dealt with uncertainty in belief modeling, and processing situated dialogue [38]. Janíček (Annex 2.3) presents a new model of the abductive continual approach to situated dialogue processing. This model extends previous work by dealing with partial proofs. Partial proofs are conditioned on the verification of explicitly represented knowledge gaps. This makes it possible to deal with both uncertainty and incompleteness in situated dialogue processing.*

Furthermore, we have worked on conversions of the CCG grammar used for parsing and realization, to other formats. The main intention here is to transform (or approximate) the CCG grammar automatically into a grammar that can be used directly by speech recognition.

Transformation of CCG grammars *Within a dialogue system, several processes typically need grammar resources, e.g. speech recognition, parsing, realization. It is not uncommon for a system to use different resources for these processes – which can lead however to inconsistencies in coverage. In the CogX dialogue system we use a CCG for both parsing and realization. Krieger & Kiefer (Annex 2.4) describe an approach for transforming this CCG grammar into a Typed Feature-Structure (TFS) grammar. From this grammar we already* *have techniques available to derive a context-free grammar for speech recognition (using [37]). In the end, this requires us to construct and maintain only a single grammar in the dialogue system.*

1.3 Relation to state-of-the-art

Below we situate the efforts for the main task of language learning in the context of "the" state-of-the-art. We refer the reader to the annexes for more in-depth discussions.

We consider human-robot interaction to be situated, social, and personal – particularly when we consider this interaction to involve spoken dialogue. Dialogue is typically about the world, and situating acting there [39]. It's social in that humans and robots are engaging each other, building up common understanding. [24, 9, 59]. And it's personal. How an individual understands communication is based on a personal perspective on the world, personal experience, personal abilities. In short, there is so much more to dialogue than just words. It all begins and ends with context [42].

From the viewpoint of language learning the question then is, how a robot could acquire such communicative skills. Given the focus in CogX on self-understanding and self-extension we take it that a robot *should* learn how to communicate. Whether learning is understood then to mean "from the ground up," or as in "always adapting to new situations" – if we look at learning as a life-long continuous process to deal with reality, and interacting in ever-new situations, it is clear that a robot should learn. And always be learning.

That shifts the burden, then, first of all to making clear *what* it is that should be learnt. We see spoken situated dialogue in human-robot interaction as a medium for acting. In the spirit of A. Glenberg *et al* [27, 28], "dialogue is for acting." An utterance has *actionable meaning*. It makes clear *what* is being talked about, *why* someone said it, and *how* one can act on it – cf. also [69]. It is a *pattern* that emerges in a specific context, through the bi-directional interaction between different information sources involved in making sense of dialogue [68, 71]. These sources mutually influence each other to converge at an adaptive and coherent *Gestalt*-like interpretation of how to place an utterance in the social, situated context, and act on it [4, 3, 36, 48, 39]. Meaning thus is not a static representation, constructed through a fixed application of a given set of rules. Meaning is a *process*. Thus, what we need to learn is a *process model* of how these patterns arise in context – particularly, of the decisions that give rise to these patterns. Since situated, social and personal contexts vary over time, we need to take care of stability and variation in patterning. And that requires this decision process model to be a dynamic model.

This then gets us back to the issue of how to learn such a process model. The *perspective* we take is here based in theories of the cognitive development of children.[3]

---

[3] There are several reasons for doing so. For one, nature is remarkable in what it can achieve, and –despite the obvious difference in "biology" between humans and robots– can serve as a source of inspiration when it comes to establishing what plays a role in

We start from dynamic models of cognitive development, seeing communication as a form of human activity [10, 42]. This deliberately places communication in the entire ecology in which the interaction takes place, i.e. *in medias res*. It is embodied, it is a socially situated individual and joint activity, it connects the individual's activity with the world of people, places, and objects. This has several consequences. For one, the development of communicative capabilities is intimately tied to how the individual develops its other biological and cognitive capabilities [31, 20, 62, 40, 41]. We do not consider communication (or even just language) in isolation from its use, and from other capabilities that help give rise to meaning within communication. This goes against "received" (but reductionist) wisdom in linguistics that considers a highly modularized, typically static way in which different levels of linguistic structure may be connected [23, 15]. Second, we have to deal with the pervasive variability in communication – by looking at how we can analyze that variability in terms of inherent patterns of stability and order we find in that variation. Variability enables us to deal with ever-unique contexts; and stability and order in variation makes it possible for us to anticipate what others are likely to do. What we are are looking for here is functional structure, for processes and the patterns they give rise to that explain the dynamics of communication as a system that adapts to context, and develops over time [20, 19].

This dynamic view on development outlines the basic mindset within which we then consider what plays a role in acquiring language, what kinds of functional structure this may yield, and how these structures might then be deployed during processing.

We adopt a theoretical perspective on language acquisition that extends this dynamic, ecological view to how children presumably learn the meaning of expressions. Under this perspective, language acquisition is considered to be a process that intertwines the development of linguistic, perceptual, and conceptual capabilities [7, 74, 8]. It combines two recent metaphors that have arisen in work on early cognitive development, namely the "child-as-data-analyst" metaphor from associative learning accounts, and the "child-as-theorist" metaphor that ties word learning to conceptual knowledge. The "child-as-data-analyst" metaphor captures how human infants are capable of attending to statistical regularities in sensory and perceptual input, to form an associative link between the current (embodied) experience and words [60, 61], whereas the "child-as-theorist" looks at how the acquisition of word meaning is mediated by early categorical structures which children

---

understanding communication, what plays a role in building up such understanding. If we are to build robots that understand humans, and can make themselves understand to humans, robots need to understand what makes humans understand. In other words, we need to bridge the gap between how robots view the world and could capture it in communication, and how humans do so. Furthermore, it is remarkably humbling to compare what children *can* do, to what we are able to achieve with robots.

acquire [73]. Consequently, we consider word learning to entail *reference* [73]. The social interaction between a child and another human builds up a context, based in the attention, trust and intentionality of (or rather, attributed to) the speaker [25]. Within this context, expressions used for naming things can take on a meaning that is based in a categorical structure which is more abstract than the current perception [44, 45].

We thus see a complex interplay between intention and attention, intension, and extension; cf. also [50, 69]. The social dimension of the interaction helps scaffold an understanding of what kind of meaning is to be construed. Linguistic structure helps establish the function within this meaning is placed, thus allowing for words their conceptual meanings to vary across contexts rather than being unitary constructs [73]. And the situatedness provides means for grounding these expectations. This moves beyond an associative account of meaning in that this constructive form of meaning can be used to acquire words for concepts that have no real-world counterparts, or where no observation is (currently) possible [5]. This interplay requires an interaction between different processes that is *interparticipatory* [20]: Processes mutually influence each other. They participate in each other's functioning like a "glass-box model," unlike a "black-box model" [39]. We can base this type of processing in multisensory processing accounts of situated language, in line with the cross-modal view of CogX: Different processes mutually influence each other to converge at a joint interpretation, reducing the overall uncertainty and incompleteness in understanding, cf. [12, 34, 33, 52].

However, at this point we are still at a fairly abstract level of characterizing how communication may be learnt. Learning involves bringing into play social, situated, and personal context. Meaning gets construed through an interaction of these different contexts, an interaction which we can trace back to models of multi-sensory or cross-modal processing. Yet that still does not immediately provide us with answers to several important questions, namely how to deal with *referential uncertainty*, and with *construal uncertainty*. Referential uncertainty points to the problem of how to determine what aspect of a situated context to link a word to, as a child may perceive many more aspects that are unrelated to an expression; cf. [51, 26, 49]. Construal uncertainty concerns the difficulty in constructing the relational structure to which an expression is to be linked, for example spatial or aspectual structure [5]. Children appear to deal with this through a combination of social context and reconstruction of a speaker's attention and intentions ("theory of mind", cf. [7, 25]), together with their own embodied experience and categorization of the environment. We try to give a concrete structure to this process, as follows. The experience itself raises, through its possible groundings in a categorical system, beliefs and expectations about possible ways of structuring the environment (cf. e.g. [6]) – or indicates where there are possibly gaps in understanding it (as for example in fast mapping [13, 14, 7, 1]). As the child hears an utterance unfold, these purely private beliefs and expectations get compared to those beliefs and expectations that the child can construct from the utterance – i.e. to a reconstruction of the categorical perspective of the "other," grounded spatiotemporally in the situation given what the other and the child (can) attend to. The way expressions are formulated into an utterance helps structure this (re)construction, but it is the intentions and conceivable intensions that mediate establishing the grounded and anticipated meanings of these expressions. This is witnessed by a wide variety of psycholinguistic insights in how (adult) humans process visually situated dialogue [4, 3, 48, 32, 36, 70]. This gradually leads us to a formulation of meaning "representation" that has its basis in social sign processes, dating back to classical American philosophy (C.S. Peirce, G.H. Mead) and functionalist linguistics (R. Jakobson [35]).

Next, we make the step from human-based theory to computing-based practice. Our goal is after all to develop an approach for a robot to learn language, communicative skills. We need to contrast the notion of meaning we arrived at above, and the way it may be acquired, with computational models. Current approaches to unsupervised learning of grammatical structure, and verb structure, all display a strong tendency to base such learning on distributional models, (alike associative models in language acquisition); cf. [56, 57]. Any referential form of context is typically left out – until learning becomes based in situated dialogue. Then, approaches use categorical labels to provide a context [2], or robot perception [54] possibly mediated by categorical structure [65, 66]. Approaches in robot language learning thereby often refer to a notion of sign, to structure the relation between perception and categorical interpretation [55, 65]. We argue that this relation however completely misses a notion of social and intentional context, a sense of *use* – thus, that it is fundamentally different from the perspective we adopt here.

It is against this complex background that we then formulate our approaches. Greeve (Annex 2.2) adopts the Language Games of Steels *et al* [63, 64, 65, 66], to investigate on how spatial language could be acquired using a notion of meaning-based self-extension grounded in Fluid Construction Grammar [67]. Kruijff (Annex 2.1) presents a novel approach to language learning, adopting Seginer's notion of common cover links to construct multi-modal sign structures. These sign structures capture linguistic structure, situationally grounded referential meaning, and the (intentional) impact of such meaning on the belief states of agents. These signs can be acquired incrementally through interaction with the world and other agents. An understanding of an utterance is based on the dynamic interplay between expected and observed structure and meaning, and an (adaptive) decision process to guide the actual composition of understanding across the linguistic, situated, and intentional dimensions. The guiding principles behind this approach are a functional Principle of Contrast, which states that any differentiation in meaning requires a difference in function [16, 46, 58, 47]. The Principle of Common Ground states that the purpose of communication is to build up and maintain a mutual understanding of what is under discussion (i.e. "common ground," [17]). This principle establishes the need for a process of differentiating and aligning the purely private perspective of the robot, with the communicated content of the human speaker. And within this process, the Principle of Contrast guides the reasoning for said differentiation and alignment. It is then the third and final principle that drives the robot to *act* upon this: the Principle of Commitment. This principle states that those involved in a dialogue are committed to the communicated content. This drives why a robot would try to make the step from private to mutual understanding, and why it would seek clarification if it were unsure. Together, these principles motivate the how, what, and why behind processing communication. Kruijff (Annex 2.1) discusses how a robot can acquire communicative skills which cover expressions down to the level of individual words, meanings, and communicative use of assertions and questions through "testimonial" interactions concerned with naming objects [72, 25] as we find them for example in the CogX George system, all in an integrated fashion. Following [36], the resulting process of meaning composition displays important cognitive properties such as being incremental, anticipatory, integrative, adaptive, and coordinated. This sets this approach apart from many other approaches which focus purely on learning structure for a single dimension, or limit themselves to only a subset of the afore properties; cf. also [43].

The contributions we aim to make are the following (particularly, cf. Annex 2.1):

- **The approach explicitly integrates *usage* into language learning, to create interpretations as *actionable meaning***: The mapping between linguistic structure and categorical interpretations of situated observations is mediated by how a (projected) mapping would affect the construction of believes, upon which the robot is to act. This goes beyond Pinker's hypothesis [49, 18] in considering variance in contexts, and as such variance in the functions and meanings of expressions (cf. [16]).

  Grammar induction methods only focus on learning syntactic structure. In human-robot interaction, learning is either unmediated [54], or guided [53] or mediated by categorical structure [65, 66]. Communicative usage is typically not considered, though a connection with action-based intentions is occasionally made [29, 30, 21, 22].

- **The approach makes no assumption as to what linguistic structure such actionable meaning can be predicated to**: Linguistic structure is acquired and composed purely on functional grounds, not on formal grounds (e.g. parts of speech). In human-robot interaction, learning focuses primarily on a small subset of expressions, notably nominal expressions or prepositional phrases; (though see related work such as [22, 55]). The approach we discuss here brings recent insights from grammar induction [57] to human-robot interaction.

2 Annexes

2.1 Kruijff, "How a Robot Could Learn From Others How To Talk" (report)

Bibliography G.J.M. Kruijff. "How a Robot Could Learn From Others How To Talk." Technical report. July 2011.

Abstract Dialogue is a situated, social activity. Meaning communicated in dialogue reflects this. The article presents **work in progress** on a novel approach to representing, learning, and constructing such meaning. The approach is based on a notion of *sign*, and *actionable meaning.* A sign is the basic representation built up in the process of understanding linguistic expressions and their relation to situated observations. This relation is inherently mediated by categorical knowledge, and it is bi-directional in that both dimensions can project and use each other's structure in guiding local interpretation. From this mediation, possible interpretations result which function as the basis for the formation of beliefs. Here, the paper makes a distinction between the beliefs of the speaker (the "other") and those of the hearer (the "I"). It is in the contrast of these beliefs, the former construed from the communicated information, the latter from personal experience of the environment, that drives how private understanding might get updated to make it possible for common ground to arise – or not. This is what is meant by actionable meaning – the construction of situated interpretations of language, driven by an understanding of the intentionality behind communication as a social activity, in terms of its impact on the formation of common ground.

The approach is grounded in a perspective on cognitive development that deliberately places communication in the entire ecology in which the interaction takes place. This extends to the acquisition of language. Acquisition is considered as a process that closely couples the development of linguistic knowledge with the development of knowledge and capabilities in other domains, e.g. categorical knowledge. The article traces this perspective through theories on child language acquisition, and how (adult) humans process visual language. The article contrasts this with current work on computational models of language learning, which are either non-situated (grammar induction, most work on verb frame induction) or lack a notion of intentionality in interaction (most work on symbol grounding in human-robot interaction). The article then presents the theoretical basis of the approach, arguing how an integrated framework can be provided for modeling language learning set in a situated, social, personal context.

Relation to WP The report describes a novel approach for a robot to acquire communicative skills, in a situated, and social, setting (Task 6.4).

Acquisition encompasses form, function, and use: The *form* of expressions, their *function* in describing experience, and their *use* in conveying beliefs and intentions.

2.2 Greeve, "Modelling the Acquisition of Grammatical Categories for Spatial Relations in Dutch" (MSc thesis)

Bibliography F. Greeve, "Modelling the Acquisition of Grammatical Categories for Spatial Relations in Dutch." MSc thesis. Department of Computational Linguistics, University of the Saarland. July 2011.

Abstract From studies on child language acquisition a general outline has emerged on the development of verbal competence of children. However, the principles which underly this process cannot be measured or tested on children. Computational models can be used to simulate the acquisition process, so hypotheses on the learning strategies of children can be explored. The focus of this thesis is on the acquisition process of spatial language.

Studies in child language acquisition stress the interplay between the perceptual, conceptual, and the linguistic system in this process. In this thesis the importance of the relation of the systems in spatial language is tested, using the Dutch language as a test case.

Dutch uses different distribution of adjectives, adverbs and prepositions to express different spatial relations. A language game is used to model the learning process, where Fluid Constructional Grammar is used to formalise the language to be acquired.

The model presented in this thesis shows how mapping spatial concepts to grammatical positions is enough to learn the difference between syntactic categories.

Relation to WP The report describes an approach for a robot to acquire grammatical knowledge for understanding and producing spatial expressions in Dutch, using a Language Games paradigm (Task 6.4). Learning encompasses the dimensions of form and meaning, formulated in terms of constructions of Fluid Construction Grammar.

2.3 Janíček, "Abductive Reasoning for Continual Dialogue Understanding" (ESSLLI SS 2011)

Bibliography M. Janíček, "Abductive Reasoning for Continual Dialogue Understanding." Proceedings of the ESSLLI Student Session 2011. Ljubljana, Slovenia. August 2011.

Abstract In this paper I present a continual context-sensitive abductive framework for understanding situated spoken natural dialogue. The framework builds up and refines a set of partial defeasible explanations of the spoken input, trying to infer the speakers intention. These partial explanations are conditioned on the eventual verification of the knowledge gaps they contain. This verification is done by executing test actions, thereby going beyond the initial context. The approach is illustrated by an example set in the context of human-robot interaction.

Relation to WP The paper describes a novel approach for situated dialogue processing to deal with incompleteness in understanding, through the construction of partial proofs.

2.4 Krieger and Kiefer, "Converting CCGs into Typed Feature Structure Grammars" (HPSG 2011)

Bibliography H.U. Krieger and B. Kiefer, "Converting CCGs into Typed Feature Structure Grammars." In: Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar. Seattle, Washington. August 2011.

Abstract In this paper, we report on a transformation scheme that turns a Categorial Grammar (CG), more specifically, a Combinatory Categorial Grammar (CCG; see (Baldridge 2002)) into a derivation- and meaning-preserving typed feature structure (TFS) grammar. We describe the main idea which can be traced back at least to work by (Karttunen 1986), (Uszkoreit 1986), (Bouma 1988), and (Calder, Klein, & Zeevat 1988). We then show how a typed representation of complex categories can be extended by other constraints, such as modes, and indicate how the Lambda semantics of combinators is mapped into a TFS representation, using unification to perform ?-conversion and ?-reduction (Barendregt 1984). We also present first findings concerning runtime measurements, showing that the PET system, originally developed for the HPSG grammar framework, outperforms the OpenCCG parser by a factor of more than 10.

Relation to WP The paper describes an essential step in synchronizing the grammatical resources used in parsing and realization, with those used for speech recognition.

References

[1] A. Alishahi, A. Fazly, and S. Stevenson. Fast mapping in word learning: What probabilities tell us. In *Proceedings of the 12th Conference on Computational Language Learning (CoNLL 2008)*, 2008.

[2] A. Alishahi and S. Stevenson. A computational model of learning semantic roles from child-directed language. *Language and Cognitive Processes*, 25:50–93, 2009.

[3] G.T.M. Altmann and Y. Kamide. Now you see it, now you don't: Mediating the mapping between language and the visual world. In J.M. Henderson and F. Ferreira, editors, *The Interface of Language, Vision, and Action: Eye Movements and The Visual World*, pages 347–386. Psychology Press, New York NY, 2004.

[4] G.T.M. Altmann and M. Steedman. Interaction with context during human sentence processing. *Cognition*, 30(3):191–238, 1988.

[5] S. Arunachalam and S.R. Waxman. Language and conceptual development. *Wiley Interdisciplinary Reviews: Cognitive Science*, ???:???, 2010.

[6] L.W. Barsalou. Perceptual symbol systems. *Behavioral & Brain Sciences*, 22:577–660, 1999.

[7] P. Bloom. *How Children Learn the Meanings of Words*. The MIT Press, 2000.

[8] A.E. Booth and S.R. Waxman. Taking stock as theories of word learning take shape. *Developmental Science*, 11(2):185–194, 2008.

[9] C. Breazeal. *Designing Sociable Robots*. The MIT Press, 2002.

[10] U. Bronfenbrenner and P.A. Morris. The ecology of developmental processes. In R. M. Lerner, editor, *Theoretical models of human development*, volume 1 of *Handbook of Child Psychology (5th Ed.)*, pages 993–1028. Wiley, Hoboken, NJ, 1998.

[11] R.A. Brooks. *Cambrian Intelligence: The Early History of the New AI*. The MIT Press, 1999.

[12] G. Calvert. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex*, 11:1110–1123, 2001.

[13] S. Carey. The child as word-learner. In M. Halle, J. Bresnan, and G.A. Miller, editors, *Linguistic theory and psychological reality*. The MIT Press, 1978.

[14] S. Carey and E. Bartlett. Acquiring a single new word. *Papers and Reports on Child Language Development*, 15:17–29, 1978.

[15] N.A. Chomsky. *Knowledge of language: Its nature, origin, and use.* Praeger, 1986.

[16] E.V. Clark. The principle of contrast: A constraint on language acquisition. In B. MacWhinney, editor, *Mechanisms of Language Acquisition: The twentieth annual Carnegie Symposium on Cognition*, pages 1–33. Erlbaum, 1987.

[17] H. Clark. *Using Language.* Cambridge University Press, 1996.

[18] A. Fazly, A. Alishahi, and S. Stevenson. A probabilistic computational model of cross-situational word learning. *Cognitive Science*, 34:1017–1063, 2010.

[19] K. W. Fischer and Z. Yan. The development of dynamic skill theory. In R. Lickliter and D. Lewkowicz, editors, *Conceptions of development: Lessons from the laboratory*, pages 279–312. Psychology Press, 2002.

[20] K.W.. Fischer and T.R. Bidell. Dynamic development of psychological structures in action and thought. In R. M. Lerner, editor, *Theoretical models of human development*, volume 1 of *Handbook of Child Psychology (5th Ed.)*, pages 467–562. Wiley, Hoboken, NJ, 1998.

[21] M. Fleischman and D. Roy. Representing intentions in a cognitive model of language acquisition: Effects of phrase structure on situated verb learning. In *AAAI Spring Symposia on Intentions in Intelligent Systems*, 2007.

[22] M. Fleischman and D. Roy. Grounded language modeling for automatic speech recognition of sports video. In *Proceedings of ACL-08: HLT*, pages 121–129, Columbus, OH, 2008.

[23] J.A. Fodor. *The Modularity of Mind.* The MIT Press, Cambridge MA, 1983.

[24] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42:143–166, 2002.

[25] S.A. Gelman. Learning from others: Children's construction of concepts. *Annual Review of Psychology*, 60:115–140, 2009.

[26] L. Gleitman. The structural sources of verb meanings. *Language Acquisition*, 1:135–176, 1990.

[27] A.M. Glenberg. What memory is for. *Behavioral & Brain Sciences*, 20:1–55, 1997.

[28] A.M. Glenberg and M.P. Kaschak. Grounding language in action. *Psychonomic Bulletin & Review*, 9(3):558–565, 2002.

[29] P. Gorniak and D. Roy. Probabilistic grounding of situated speech using plan recognition and reference resolution. In *Proceedings of the Seventh International Conference on Multimodal Interfaces (ICMI 2005)*, 2005.

[30] P. Gorniak and D. Roy. Situated language understanding as filtering perceived affordances. *Cognitive Science*, 31(2):197–231, 2007.

[31] G. Gottlieb, D. Wahlsten, and R. Lickliter. The significance of biology for human development: A developmental psychobiological systems view. In R. M. Lerner, editor, *Theoretical models of human development*, volume 1 of *Handbook of Child Psychology (5th Ed.)*, pages 233–274. Wiley, Hoboken, NJ, 1998.

[32] K. Hadelich and M.W. Crocker. Gaze alignment of interlocutors in conversational dialogues. In *Proceedings of the 19th CUNY Conference on Human Sentence Processing*, New York, USA, 2006.

[33] G. Hickok. The functional neuroanatomy of language. *Physics of Life Reviews*, 6:121–143, 2009.

[34] G. Hickok and D. Poeppel. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 4(4):67–99, 2004.

[35] R. Jakobson. *On Language.* Harvard University Press, 1990. L.R. Waugh and M. Monville-Burston (Eds.).

[36] P. Knoeferle and M.C. Crocker. The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cognitive Science*, pages 481–529, 2006.

[37] H.U. Krieger. From UBGs to CFGs: A practical corpus-driven approach. *Natural Language Engineering*, 13(4):317–351, 2006.

[38] G.J.M. Kruijff, M. Janíček, and P. Lison. Continual processing of situated dialogue in human-robot collaborative activities. In *Proceedings of the 19th International Symposium on Robot and Human Interactive Communication (RO-MAN 2010).* 2010.

[39] G.J.M. Kruijff, P. Lison, T. Benjamin, H. Jacobsson, H. Zender, and I. Kruijff-Korbayová. Situated dialogue processing for human-robot interaction. In H.I. Christensen, G.J.M. Kruijff, and J. L. Wyatt, editors, *Cognitive Systems*, volume 8 of *Cognitive Systems Monographs, COSMOS*, chapter 8, pages 311–364. Springer Verlag, Berlin/Heidelberg, Germany, 2010.

[40] G. Lakoff. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* The University of Chicago Press, 1987.

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

[42] D. Magnusson and H. Stattin. The person in the environment: Towards a general model for scientific inquiry. In R. M. Lerner, editor, *Theoretical models of human development*, volume 1 of *Handbook of Child Psychology (5th Ed.)*, pages 685–760. Wiley, Hoboken, NJ, 1998.

[43] M. Mayberry, M.W. Crocker, and P. Knoeferle. Learning to attend: A connectionist model of the coordinated interplay of utterance, visual context, and world knowledge. *Cognitive Science*, 33:449–496, 2009.

[44] L.L. Namy and S.R. Waxman. Naming and exclaiming: Infants' sensitivity to naming contexts. *Journal of Cognitive Development*, 1:405–428, 2000.

[45] L.L. Namy and S.R. Waxman. Symbols redefined. In L.L. Namy, editor, *Symbol Use and Symbolic Representation: Developmental and Comparative Perspectives*, pages 269–277. Lawrence Erlbaum Associates, 2005.

[46] J. Panevová. On verbal frames in functional generative description I. *Prague Bulletin of Mathematical Linguistics*, 22:3–40, 1974.

[47] J. Panevová. Valency frames and the meaning of the sentence. In P. Luelsdorff, editor, *The Prague School of Structural and Functional Linguistics*, volume 41 of *Linguistics and Literary Studies in Eastern Europe*, pages 223–243. John Benjamins, Amsterdam, 1994.

[48] M.J. Pickering and S. Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–225, 2004.

[49] S. Pinker. *Learnability and cognition: The acquisition of argument structure.* The MIT Press, 1989.

[50] H. Putnam. Meaning and reference. *Journal of Philosophy*, 70:699–711, 1973.

[51] W.V.O. Quine. *Word and Object.* The MIT Press, 1960.

[52] C.W. Robinson and V.M. Sloutsky. Development of cross-modal processing. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1:135–141, January/February 2010.

[53] D. Roy and N. Mukherjee. Towards situated speech understanding: visual context priming of language models. *Computer Speech & Language*, 19(2):227–248, April 2005.

[54] D.K. Roy. Learning words and syntax for a scene description task. *Computer Speech and Language*, 16(3), 2002.

[55] D.K. Roy. Semiotic schemas: A framework for grounding language in action and perception. *Artificial Intelligence*, 167(1-2):170–205, 2005.

[56] S. Schulte im Walde. The induction of verb frames and verb classes from corpora. In A. Lüdeling and M. Kytö, editors, *Corpus Linguistics: An International Handbook*, volume 2, pages 952–972. Walter de Gruyter, 2009.

[57] Y. Seginer. *Learning Syntactic Structure.* PhD thesis, Institute of Logic, Language and Information, University of Amsterdam, 2007.

[58] P. Sgall, E. Hajičová, and J. Panevová. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects.* Reidel, 1986.

[59] C. L. Sidner, C. Lee, C. Kidd, N. Lesh, and C. Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1-2):140–164, 2005.

[60] V.M. Sloutsky. The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, 7(6):246–251, 2003.

[61] V.M. Sloutsky, H. Kloos, and A.V. Fisher. When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science*, 18(2):179–185, 2007.

[62] E.S. Spelke and E.L. Newport. Nativism, empiricism, and the development of knowledge. In R. M. Lerner, editor, *Theoretical models of human development*, volume 1 of *Handbook of Child Psychology (5th Ed.)*, pages 275–340. Wiley, Hoboken, NJ, 1998.

[63] L. Steels. A self-organizing spatial vocabulary. *Artificial Life Journal*, 2(3):319332, 1996.

[64] L. Steels. How to do experiments in artificial language evolution and why. In A. Cangelosi, A. Smith, and J. Smith, editors, *Proceedings of the 6th International Conference on The Evolution of Language (EVOLANG6)*, 2006.

[65] L. Steels. Semiotic dynamics for embodied agents. *IEEE Intelligent Systems*, 21(3):32–38, 2006.

[66] L. Steels. The symbol grounding problem has been solved. so what's next? In M. De Vega, G. Glennberg, and G. Graesser, editors, *Symbols, embodiment and meaning.* Academic Press, New Haven, 2008.

[67] L. Steels. A first encounter with fluid construction grammar. In L. Steels, editor, *Design Patterns in Fluid Construction Grammar*. John Benjamins Publishing Company, 2011.

[68] B.E. Stein and M.A. Meredith. *The Merging of the Senses*. Cognitive Neuroscience Series. The MIT Press, 1993.

[69] M. Stone. Intention, interpretation and the computational structure of language. *Cognitive Science*, 28(5):781–809, 2004.

[70] M.K. Tanenhaus and S. Brown-Schmidt. Language processing in the real world. *Philosophical transactions of the Royal Society*, 363:1105–1122, 2008.

[71] M.T. Wallace. The development of multisensory integration. In G. Calvert, C. Spence, and B.E. Stein, editors, *The Handbook of Multisensory Processes*, pages 625–642. The MIT Press, 2004.

[72] S.R. Waxman. Early word learning and conceptual development: Everything had a name, and each name gave birth to a new thought. In U. Goswami, editor, *Blackwell Handbook of Childhood Cognitive Development*, pages 102–126. 2004.

[73] S.R. Waxman and S.A. Gelman. Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, 13(6):258–263, 2009.

[74] S.R. Waxman and J.L. Lidz. Early word learning. In D. Kuhn and R.S. Siegler, editors, *Handbook of Child Psychology*, volume 2: Cognition, Perception, and Language, pages 687–733. John Wiley, 2006.

Abductive Reasoning for Continual Dialogue Understanding

Miroslav Janíček

German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
`mjanicek@dfki.de`

Abstract. In this paper I present a continual context-sensitive abductive framework for understanding situated spoken natural dialogue. The framework builds up and refines a set of partial defeasible explanations of the spoken input, trying to infer the speaker's intention. These partial explanations are conditioned on the eventual verification of the knowledge gaps they contain. This verification is done by executing test actions, thereby going beyond the initial context. The approach is illustrated by an example set in the context of human-robot interaction.

1 Introduction

In task-oriented dialogues between two agents, such as between two humans or a human and a robot, there is more to dialogue than just understanding words. The agents need to understand what is being talked about, but it also needs to understand why it was told something. In other words, what the speaker *intends* the hearer to do with the information in the larger context of their joint activity.

Therefore, understanding language can be phrased as an *intention recognition* problem: given an utterance from the human, how do we find the intention behind it?

In this paper, I explore an idea inspired by the field of continual planning [4], by explicitly capturing the possible knowledge gaps in such an interpretation. The idea is based on the notion of *assertion*, an explicit test for the validity of a certain fact, going beyong the current context.

The structure of the paper is as follows. After briefly introducing the notion of intention recognition and abduction in the next section, I introduce the continual abductive reasoning mechanism in §3, and discuss it on an example in §4, before concluding with a short summary.

2 Background

The idea of expressing *understanding* in terms of intention recognition has been introduced by H. P. Grice [7,14]. In this paper, I build on Stone and Thomason's approach to the problem [17] who in turn extend the work done by Hobbs and others [8], and base their approach to intention recognition on *abductive reasoning.*

Abduction. Abduction is a method of explanatory logical reasoning introduced into modern logic by Charles Sanders Peirce [6]. Given a theory T, a rule $T \vdash A \rightarrow B$ and a fact B, abduction allows inferring A as an explanation of B. B can be deductively inferred from $A \cup T$. If $T \nvdash A$, then we say that A is an *assumption.*

There may be many possible causes of B besides A. Abduction amounts to *guessing*; assuming that the premise is true, the conclusion holds too. To give a well-known example:

> Suppose we are given two rules saying "if the sprinkler is on, then the lawn is wet" and "if it rained, then the lawn is wet". Abductively inferring the causes for the fact that the lawn is wet then yields two possible explanations: the sprinkler is on, or it rained.

Obviously, as there may be many possible explanations for a fact, in practical applications there needs to be a mechanism for selecting the best one. This may be done by purely syntactic means (e.g. lengths of proofs), or semantically by assigning *weights* to abductive proofs and selecting either the least or most costly proof [16], or by assigning probabilities to proofs [12]. In that case, the most probable proof is also assumed to be the best explanation. Our approach combines both aspects.

Intention recognition. Abduction is a suitable mechanism to perform inferences on the pragmatic (discourse) level. For understanding, abduction can be used to infer the explanation *why* an agent said something, in other words the *intention* behind the utterance. Reversing the task, given an intention, we may infer the way *how* to express it [18]. Intentions can therefore serve as a middle representational layer and abduction as the inference mechanism by using which we either turn a realisation into an intention, or the other way around.

3 Approach

This paper extends the work of Stone and Thomason on context-sensitive language understanding by explicitly modelling the knowledge gaps that inevitably arise in such an effort due to uncertainty and partial observability. The approach is based on generating partial hypotheses for the explanation of the observed behaviour of other agents, under the assumption that the observed behaviour is intentional. These partial hypotheses are defeasible and conditioned on the validity (and eventual verification) of their assumptions.

In this section, I examine the an abductive reasoning system capable of representing knowledge gaps in the form of partial proofs, how such partial proofs can be generated and verified or falsified, and the semantic framework used in our system to capture linguistic meaning that the system then grounds in reality.

3.1 Partial Abductive Proofs

Our abductive inference mechanism is essentially Hobbs and Stickel's logic programming approach to weighted abduction [8,16] enhanced by a contextual aspect [1] with the weights in the system being assigned a probabilistic interpretation following Charniak and Shimony [5].

Proof procedure. Formally, inference in our system makes use of four ingredients: *facts*, *rules*, *disjoint declarations* and *assumability functions*, collectively called the *abduction context*, and using these iteratively in order to derive proofs of an initial *goal*.

- Facts are modalised formulas of the form

  $$\mu : A$$

  where μ is a (possibly empty) sequence of contexts, and A is an atomic formula, possibly containing variables.
- Rules are modalised Horn clauses, i.e. formulas of the form

  $$(\mu_1 : A_1/t_1) \wedge ... \wedge (\mu_n : A_n/t_n) \rightarrow (\mu_H : H)$$

  where each of the $\mu_i : A_i$ and $\mu_H : H$ are modalised formulas. Each antecedent is annotated by t_i, which determines the way the antecedent is manipulated and is one of the following:
  - *true* – the antecedent has to be proven, i.e. either it is a fact, or a head of some rule;
  - *assumable(f)* – the antecedent is assumable under function f;
  - *assertion* – the antecedent is asserted, i.e. identifies a knowledge gap, conditioning the validity of the proof on it being proved in a subsequent reinterpretation (see below).
- Assumability functions are partial functions f, $f : \mathcal{F} \rightarrow \mathbb{R}_0^+$, where $\mathcal{F}$ is the set of modalised formulas, with the additional monotonicity property that if $F \in \mathrm{dom}(f)$, then for all more specific (in terms of variable substitution) facts F', $F' \in \mathrm{dom}(f)$ and $f(F) \leq f(F')$.
- A disjoint declaration is a statement of the form

  $$\mathrm{disjoint}([\mu : A_1, ..., \mu : A_n])$$

  which specifies that at most one of the modalised formulas $\mu : A_i$ may be used in the proof. A_i and A_j cannot be unified for all $i \neq j$.

A *proof state* is a sequence of marked modalised formulas (called *queries* in this context)

$$Q_1[n_1], ...Q_m[n_m]$$

The markings n_i are one of the following:

Algorithm 1 (Nondeterministic) weighted abduction

```
$c$ = the abduction context
$L$ = the initial proof state

while $L$ contains a query marked as unsolved:
    $Q \leftarrow$ leftmost query in $L$ marked as unsolved
    choose a transformation rule $t$ so that APPLY-RULE$(t, Q, L, c)$ succeeds
    $L \leftarrow$ APPLY-RULE$(t, Q, L, c)$
```

- *unsolved(f)* – the query is yet to be proved, assumable under assumability
- *proved* – the query is proved or in the process of being proved;
- *assumed(f)* – the query is assumed under f;
- *asserted* – the query is asserted

The proof procedure starts from a single query marked as unsolved (called the *goal*), iteratively rewriting the proof state by manipulating the leftmost unsolved query Q_l. First, the query has to pass constraints imposed by disjoint declarations. If it does, it is either proved (using facts or rules), assumed under an assumability function, or eliminated if any of the queries to the right is unifiable with Q_l. In other words, each query is proved or assumed at most once.

The initial query Q is proved when there is no unsolved query in the proof state. The final proof state Π_Q is then the proof of Q. The proof procedure is schematised by Algorithm 1. Note that the proof procedure assures that the cost of the proofs are monotonic with respect to unification and application of rules, allowing for the use of efficient search strategies.

Knowledge gaps and assertions. Our extension of the "classical" logic-programming-based weighted abduction as proposed by Stickel and Hobbs lies in the extension of the proof procedure with the notion of *assertion* based on the work in continual automated planning [4], allowing the system to reason about information not present in the knowledge base, thereby addressing the need for reasoning under the open-world assumption.

In continual automated planning, assertions allow a planner to reason about information that is not known at the time of planning (for instance, planning for information gathering), an assertion is a construct specifiying a "promise" that the information in question will be resolved eventually.

By using a logic programming approach, we can use unbound variables in the asserted facts in order to reason not only about the fact that the given assertion will be a fact, but also under-specify its eventual arguments.

The proposed notion of an *assertion* for our abductive system is based on *test actions* $\langle F \rangle$ [2]. Baldoni et al. specify a test as a proof rule. In this rule, a goal F follows from a state $a_1, ..., a_n$ after steps $\langle F \rangle, p_1, ..., p_m$ if we can establish F on $a_1, ..., a_n$ with answer σ and this (also) holds in the final state resulting from executing $p_1, ..., p_m$.

An assertion is the transformation of a test into a partial proof which assumes the verification of the test, while at the same time conditioning the obtainability of the proof goal on the tested statements. $\mu : \langle D \rangle$ within a proof $\Pi[\langle D \rangle]$ to a goal C turns into $\Pi[D] \rightarrow C \wedge \mu : D$. Should $\mu : D$ not be verifiable, Π is invalidated.

Probabilistic interpretation. In weighted abduction, weights assigned to assumed queries are used to calculate the overall proof cost. The proof with the lowest cost is the best explanation. However, weights are usually not assigned any semantics, and often a significant effort by the writer of the rule set is required to achieve expected results [8].

However, Charniak and Shimony [5] showed that by setting weights to $-\log$ of the prior probability of the query, the resulting proofs can be given probabilistic semantics.

Suppose that query Q_k can be assumed true with some probability $P(Q_k \text{ is true})$. Then if Q_k is assumable under assumability function f such that $f(Q_k) = -\log(P(Q_k \text{ is true}))$, and under the independence assumption, we can represent the overall probability of the proof $\Pi = Q_1[t_1], ..., Q_n[t_n]$ as

$$P(\Pi) = e^{\sum_{k=1}^{n} cost(Q_k)}$$

where

$$cost(Q_k) = \begin{cases} f(Q_k) & \text{if } m_i = \mathit{assumed}(f) \\ 0 & \text{otherwise} \end{cases}$$

The best explanation Π_{best} of a query Q is then

$$\Pi_{best} = \underset{\Pi \text{ proof of } Q}{\arg\min} \; P(\Pi)$$

Exact inference in such a system is NP-complete, and so is approximate inference given a threshold [5]. However, it is straightforward to give an anytime version of the algorithm – simply by performing iterative deepening depth-first search [13] and memorizing a list of most probable proofs found so far.

3.2 Generating Partial Hypotheses

For each goal G, a determinisation of Algorithm 1 returns a set of proofs H, with a total ordering on this set. Due to the use of assertions, some of these proofs may be partial, and their validity has to be verified. The presence of assertions in the proofs means that there is a knowledge gap, namely the truth value of the assertion. Each assertion thus specifies the need for performing a (test) action. This action might require the access to other knowledge bases than the abductive context, as in the case of resolving referring expressions, or an execution of a physical action.

Formally, given an initial goal G and context c, the abduction procedure produces a set H of hypotheses $c : \Pi \rightarrow C \wedge c_i : A_i$, where c_i is a sub-context in

Algorithm 2 (Nondeterministic) continual abduction

```
CONTINUAL-ABDUCTION($c, \Pi$):
  $c$ = context
  $\Pi$ = proof

  while $\Pi$ contains assertion $A$:
    $c' \leftarrow$ TEST-ACTION($c, A$)
    $H \leftarrow$ ABDUCE($c', A$)
    for all $\Pi' \in H$:
      CONTINUAL-ABDUCTION($c', \Pi'$)
  return $\Pi$
```

which where an assertion $A_i \in \Pi$ may be evaluated. Such proofs are thus both *partial* and *defeasible* — they may be both extended and discarded, depending on the evaluation of the assertions.

The set of possible hypotheses is continuously expanded until the best full proof is found. This process is defined in Algorithm 2.

The algorithm defines the search space in which it is possible to find the most probable proof of the initial goal G. The important point is, however, that it is just that — a definition. The actual implementation may keep track of the partial hypotheses it defines, and take the appropriate test actions when necessary, or postpone them indefinitely.

3.3 Representing Linguistic Meaning

For representing linguistic meaning in our system, we use the *Hybrid Logic Dependency Semantics* (HLDS), a hybrid logic [3] framework that provides the means for encoding a wide range of semantic information, including dependency relations between heads and dependents [15], tense and aspect [11], spatio-temporal structure, contextual reference, and information structure [10].

HLDS uses hybrid logic to capture dependency complexity in a model-theoretic relational structure, using ontological sorting to capture categorial aspects of linguistic meaning, and naturally capture (co-)reference by explicitly using *nominals* in the representation.

Generally speaking, HLDS represents an expression's linguistic meaning as a conjunction of modalised terms, anchored by the nominal that identifies the head's proposition:

$$@_{h:\mathsf{sort}_h}\ (\mathbf{prop}_h \wedge \langle \mathsf{R}_i \rangle\ (d_i : \mathsf{sort}_{d_i} \wedge \mathbf{dep}_i))$$

Here, the head proposition nominal is h. $\mathbf{prop}_h$ represents the *elementary predication* of the nominal h. The dependency relations (such as Agent, Patient, Subject, etc.) are modelled as modal relations $\langle R_i \rangle$, with the dependent being identified by a nominal d_i. Features attached to a nominal (e.g. ⟨Num⟩ ⟨Quantification⟩, etc.) are specified in the same way.

Figure 1 gives an example of HLDS representation (logical form) of the sentence "Take the mug". The logical form has three nominals, $event_1$, $agent_1$ and $thing_1$ that form a dependency structure: $event_1$ is the the head of dependency relations Actor (the dependent being $agent_1$), $\mathsf{Patient}$ ($thing_1$), and $\mathsf{Subject}$ ($agent_1$). Each nominal has an ontological sort (illustrated on $event_1$, the sort is $\mathsf{action\text{-}non\text{-}motion}$) a proposition (**take**), and features (Mood).

$$
\begin{aligned}
@_{event_1:\mathsf{action\text{-}non\text{-}motion}}(&\mathbf{take} \wedge \\
&\langle \mathsf{Mood} \rangle\ \mathbf{imp} \wedge \\
&\langle \mathsf{Actor} \rangle\ agent_1 : \mathsf{entity} \wedge \\
&\langle \mathsf{Patient} \rangle\ thing_1 : \mathsf{thing} \wedge \mathbf{mug} \wedge \\
&\quad \langle \mathsf{Delimitation} \rangle\ \mathbf{unique} \wedge \\
&\quad \langle \mathsf{Num} \rangle\ \mathbf{sg} \wedge \\
&\quad \langle \mathsf{Quantification} \rangle\ \mathbf{specific})) \wedge \\
&\langle \mathsf{Subject} \rangle\ (agent_1 : \mathsf{entity} \wedge \mathbf{addressee}))
\end{aligned}
$$

Fig. 1. HLDS semantics for the utterance "Take the mug"

$$
\begin{aligned}
&\mathrm{sort}(\mathrm{event}_1, \mathrm{action\text{-}non\text{-}motion}), \\
&\mathrm{prop}(\mathrm{event}_1, \mathrm{take}), \\
&\mathrm{feat}(\mathrm{event}_1, \mathrm{mood}, \mathrm{imp}), \\
&\mathrm{rel}(\mathrm{event}_1, \mathrm{actor}, \mathrm{agent}_1), \\
&\mathrm{sort}(\mathrm{agent}_1, \mathrm{entity}), \\
&\mathrm{prop}(\mathrm{agent}_1, \mathrm{addressee}), \\
&\mathrm{rel}(\mathrm{event}_1, \mathrm{patient}, \mathrm{thing}_1), \\
&\mathrm{feat}(\mathrm{thing}_1, \mathrm{delimitation}, \mathrm{unique}), \\
&\mathrm{feat}(\mathrm{thing}_1, \mathrm{num}, \mathrm{sg}), \\
&\mathrm{feat}(\mathrm{thing}_1, \mathrm{quantification}, \mathrm{specific})
\end{aligned}
$$

Fig. 2. The translation of the hybrid logic formula in Figure 1 into abduction facts

Every logical form in HLDS, being a formula in hybrid logic, can be decomposed into a set of facts in the abductive context corresponding to its minimal Kripke model. The resulting set of abduction facts obtained by decomposing the logical form in Figure 1 is shown by Figure 2.

HLDS only represents the meaning as derived from the linguistic realisation of the utterance and does not evaluate the state of affairs denoted by it. This sets the framework apart from semantic formalisms such as DRT [9]. The grounding in reality is partly provided by the continual abductive framework by generating and validating (or ruling out) partial abductive hypotheses as more information is added to the system.

4 Example

Let us examine the mechanism in an example. Suppose that a human user is dealing with a household robot capable of manipulating objects (picking them up, putting them down). The robot and the human are both looking at a table with a represented by the term "mug_1", and the human wants the robot to pick up the mug.

The human's utterance,

"Take the mug."

is analysed in terms of HLDS (see Figure 1), and its translation is made part of the abductive context c.

Suppose that proving the following goal in the context c

$$\mathrm{uttered}(\mathrm{human}, \mathrm{robot}, \mathrm{event}_1)$$

yields the following (best) proof, displayed with markings following §3.1:

$$\begin{array}{rll}
\mathrm{uttered}(\mathrm{human}, \mathrm{robot}, \mathrm{event}_1) & [\mathit{proved}] & (1) \\
\hline
\mathrm{prop}(\mathrm{event}_1, \mathrm{take}) & [\mathit{proved}] & (2) \\
\mathrm{intends}(\mathrm{event}_1, \mathrm{human}, I) & [\mathit{assumed}(p = 0.9)] & (3) \\
\mathrm{rel}(\mathrm{event}_1, \mathrm{patient}, \mathrm{thing}_1) & [\mathit{proved}] & (4) \\
\mathrm{refers\_to}(\mathrm{thing}_1, X) & [\mathit{asserted}] & (5) \\
\mathrm{pre}(I, \mathrm{object}(X)) & [\mathit{asserted}] & (6) \\
\mathrm{post}(I, \mathrm{state}(\mathrm{is\text{-}holding}(\mathrm{robot}, X))) & [\mathit{assumed}(p = 0.7)] & (7)
\end{array}$$

The proof is an explanation of the event (1) in terms of a partially specified intention I (3), defined by its pre- and post-condition. The precondition is the existence of an entity X (6), and the postcondition (7) is the state in which the robot is holding the entity X. The proof appeals to the logical form of the utterance (2, 4).

In the proof, atoms (3) and (7) are assumed under a constant assumability function that assigns them probability 0.9 and 0.7 respectively. This means that given our knowledge base, such a sentence expresses an intention of the human with prior probability 0.9, and that with prior probability 0.7, this intention has something to do with the robot physically taking holding some object (as opposed to uses such as "take a picture").

The atoms (5) and (6) are marked as *asserted*. These assertions identify the knowledge gaps in the interpretation – the interpretation is only valid if they are verified using test actions.

Suppose that the assertion (5) is tested first. This amounts to resolving the referring expression, headed by the nominal thing_1 in the logical form. The action is performed, giving rise to a new abduction context c', in which the abduction context is updated by specifying a reference resolution function r, yielding the following two hypotheses:

$$\begin{array}{rl}
\mathrm{refers\_to}(\mathrm{thing}_1, \mathrm{mug}_1) & [\mathit{proved}] \\
\hline
\mathrm{ref}(\mathrm{thing}_1, \mathrm{mug}_1) & [\mathit{assumed}(r)]
\end{array}$$

(i.e. p_1 refers to the mug under assumability function r),

$$\frac{\text{refers\_to}(\text{thing}_1, X)\ [\textit{proved}]}{\text{unknown-referent}(\text{thing}_1, X)\ [\textit{assumed}(p = \textit{0.4})]}$$

i.e. the reference was not resolved. This accounts for the possibility of misunderstanding, where the human might be referring to an object that is not part of common ground, and the reference thus cannot be resolved.

Now that the assertion (5) has been tested, the system can check the assertion (6). Depending on the value of the assumability function r above, it might first perform the test action for existence in the former (if $r > 0.4$), or the latter context (if $r < 0.4$), or in a random order (if $r = 0.4$).

In the former context, where the reference has been resolved to the mug, the robot might ask "Did you mean I should take *this* object?" (pointing at the mug, testing the hypothesis

$$\text{pre}(I, \text{object}(\text{mug}_1))$$

In the latter case, it might ask "Which object did you mean?", prompting the human to give an answer that would ultimately become the proof of the test action for

$$\text{pre}(I, \text{object}(X))$$

in the proof above.

5 Conclusion and Future Work

This paper presents an abductive framework for natural language understanding that is based on abductive reasoning over partial hypotheses. The framework is set within the process of intention recognition.

The abductive framework is contextually-enhanced version of a logic programming approach to weighted abduction with a probabilistic semantics assigned to the weights. Our extension of this framework is in the introduction of the notion of *assertion*, which is essentially a requirement for a future test to verify or falsify the proposition, i.e. to fill a knowledge gap about the validity of the proposition. The hypotheses are therefore defeasible in the sense that the falsification of their assertions leads to a retraction and adoption of an initially less likely alternative.

By explicitly reasoning about these knowledge gaps, the system is able to go beyond the current context and knowledge base, addressing the need for reasoning under the open-world assumption.

Future research in this area will include a more informed interface to the decision-making processes involved in the selection of the hypotheses to test, and the stability of the partial hypotheses.

6 Acknowledgments

The work reported in this paper has been supported by the EU ICT FP7 projects CogX, “Cognitive Systems that Self-Understand and Self-Extend” (project no. 215181) and NIFTi, “Natural Human-Robot Collaboration in Dynamic Environments” (project no. 247870).

I would like to thank Geert-Jan Kruijff for insightful and inspiring discussions, and the reviewers for their comments.

References

1. Baldoni, M., Giordano, L., Martelli, A.: A modal extension of logic programming: Modularity, beliefs and hypothetical reasoning. J. Log. Comp. 8(5), 597–635 (1998)
2. Baldoni, M., Giordano, L., Martelli, A., Patti, V.: A modal programming language for representing complex actions. In: Proc. of DYNAMICS’98. pp. 1–15 (1998)
3. Blackburn, P.: Representation, reasoning, and relational structures: a hybrid logic manifesto. Logic Journal of the IGPL 8(3), 339–625 (2000)
4. Brenner, M., Nebel, B.: Continual planning and acting in dynamic multiagent environments. Journal of Autonomous Agents and Multiagent Systems (2008)
5. Charniak, E., Shimony, S.E.: Probabilistic semantics for cost based abduction. In: AAAI-90 proceedings (1990)
6. Fann, K.T.: Peirce’s Theory of Abduction. Mouton, The Hague, The Netherlands (1970)
7. Grice, H.P.: Meaning. The Philosophical Review 66(3), 377–388 (1957)
8. Hobbs, J.R., Stickel, M.E., Appelt, D., Martin, P.: Interpretation as abduction. Tech. Rep. 499, AI Center, SRI International, Menlo Park, CA, USA (Dec 1990)
9. Kamp, H., Reyle, U.: From Discourse to Logic. Kluwer, Dordrecht, The Netherlands (1993)
10. Kruijff, G.J.M.: A Categorial-Modal Logical Architecture of Informativity: Dependency Grammar Logic & Information Structure. Ph.D. thesis, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic (2001)
11. Moens, M., Steedman, M.: Temporal ontology and temporal reference. Computational Linguistics 14(2), 15–28 (1988)
12. Poole, D.: Probabilistic Horn abduction and Bayesian networks. Artificial Intelligence 64(1), 81–129 (1993)
13. Russell, S.J., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall, Upper Saddle River, NJ, USA, second edn. (2003)
14. Schiffer, S.R.: Meaning. Oxford, Clarendon Press (1972)
15. Sgall, P., Hajičová, E., Panevová, J.: The Meaning of the Sentence in Its Semantic and Pragmatic Aspects. Reidel Publishing Company, Dordrecht, The Netherlands and Academia, Prague, Czechoslovakia (1986)
16. Stickel, M.E.: A Prolog-like inference system for computing minimum-cost abductive explanations in natural-language interpretation. Tech. Rep. 451, AI Center, SRI International, Menlo Park, CA, USA (Sep 1988)
17. Stone, M., Thomason, R.H.: Context in abductive interpretation. In: Proceedings of EDILOG 2002: 6th workshop on the semantics and pragmatics of dialogue (2002)
18. Stone, M., Thomason, R.H.: Coordinating understanding and generation in an abductive approach to interpretation. In: Proceedings of DIABRUCK 2003: 7th workshop on the semantics and pragmatics of dialogue (2003)

Converting CCGs into Typed Feature Structure Grammars[*]

Hans-Ulrich Krieger and Bernd Kiefer
German Research Center for Artificial Intelligence (DFKI GmbH)
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
{krieger,kiefer}@dfki.de

Introduction

In this paper, we report on a transformation scheme that turns a Categorial Grammar (CG), more specifically, a Combinatory Categorial Grammar (CCG; see (Baldridge 2002)) into a derivation- and meaning-preserving typed feature structure (TFS) grammar. We describe the main idea which can be traced back at least to work by (Karttunen 1986), (Uszkoreit 1986), (Bouma 1988), and (Calder, Klein, & Zeevat 1988). We then show how a typed representation of complex categories can be extended by other constraints, such as modes, and indicate how the Lambda semantics of combinators is mapped into a TFS representation, using unification to perform α-conversion and β-reduction (Barendregt 1984). We also present first findings concerning runtime measurements, showing that the PET system, originally developed for the HPSG grammar framework, outperforms the OpenCCG parser by a factor of more than 10.

Motivation

The Talking Robots (TR) group here at the LT Lab of DFKI uses categorial grammars in several large EU projects in order to communicate with robots in spoken language. The grammars for English and Italian are written in the OpenCCG dialect of CCG.

Faster Parser. The main rationale for our transformation method is driven by the need that we are looking for a reliable and trainable (C)CG parser that is faster than the one which comes with the OpenCCG system. People from the DFKI LT group have co-developed the PET system (Callmeier 2000), a highly-tuned TFS parser written in C++, which originally grew out of the HPSG community. In order to use such a TFS parser in a CG setting, the (combinatory) rules and lexicon entries need to be transformed into a TFS representation.

Structured Language Model. Another major rationale for the transformation comes from the fact that the CCG grammars are used for spoken language, operating on the output of a speech recognizer. Although speech recognizers are based on trained statistical models, modern recognizers can be further tuned by supplying an additional structured language model. Given a TFS grammar for the transformed CCG grammar, we would like to use the corpus-driven approximation method described in (Krieger 2007) to generate a context-free approximation of the deep grammar. This approximation then serves as our language model for the recognizer. Again, as is the case for PET, software can be reused here, since the method described in (Krieger 2007) is implemented for the external chart representation of the PET system.

Cross-Fertilization. We finally hope that our experiment provides insights on how to incorporate descriptive means from CG (e.g., direct slash notation for categories) into the HPSG framework, even though they are compiled out in the end. Thus, specification languages for HPSG, such as $\mathcal{TDL}$ (Krieger 1995), might be extended by some kind of macro formalism, allowing a grammar writer to state such extended rules. However, we will not speculate on this in the paper.

Categorial Grammar

Categorial grammar started with Bar-Hillel in 1953 who adapts and extend Ajdukiewicz's work by adding directionality to what Ajdukiewicz (by referring to Husserl) called "Bedeutungskategorie". The grammatical objects in Bar-Hillel's system are called *categories*. The set of *complex* categories C can be defined inductively by assuming a set of *atomic* categories A (e.g., s or np) and a set of binary functor symbols F_2 (usually $/$ and $\backslash$ for one-dimensional binary grammar rules):

1. *if* $\mathsf{a} \in \mathsf{A}$ *then* $\mathsf{a} \in \mathsf{C}$
2. *if* $\mathsf{c}, \mathsf{c}' \in \mathsf{C}$ *and* $\mathsf{f} \in \mathsf{F}_2$ *then* $\mathsf{cfc}' \in \mathsf{C}$

The system of categories in its simplest form is usually equipped with two very fundamental binary rules (or better, rule schemes), viz., forward (>) and backward (<) *functional application*—this is called the AB calculus (for Ajdukiewicz & Bar-Hillel). Here and in the following, we use the notation from (Baldridge 2002), originating from the work of Mark Steedman:

$$(>\mathbf{A})\quad \mathsf{X/Y}\ \ \mathsf{Y} \ \Rightarrow\ \mathsf{X}$$

---

[*] The research described here has been partly financed by the **TAKE** project (take.dfki.de), funded by the German Federal Ministry of Education and Research, and the European Integrated projects **CogX** (cogx.eu), **NIFTi** (nifti.eu), and **Aliz-e** (aliz-e.org) under contract numbers 01IW08003, FP7 ICT 215181, 247870, and 248116. We would like to thank our reviewers for their useful comments.

$(<\mathbf{A})\quad \mathsf{Y}\ \ \mathsf{X}\backslash\mathsf{Y} \ \Rightarrow\ \mathsf{X}$

Depending on the kind of slash, complex category symbols in these rules look to the right (forward) or to the left (backward) in order to derive a simpler category. Such a framework is in the truest sense *lexicalized*, since the categories in these rules are actually category schemes: there is no category X/Y, only instantiations, such as, for instance, $(\mathsf{s}\backslash\mathsf{np})/(\mathsf{s}\backslash\mathsf{np})$ for modal verbs. Furthermore, and very importantly, concrete categories are only specified for lexicon entries ($\vdash$ maps the word to its category):

defeat $\vdash (\mathsf{s}\backslash\mathsf{np})/\mathsf{np}$

Not only are lexical entries equipped with a category, but also with a semantics. Since Montague, categorial grammarians have often used the Lambda calculus to make this explicit. Abstracting away from several important things such as tense, we can define what is meant by the transitive verb *defeat* (: is used to attach the semantic to a lexicon entry):

defeat $\vdash (\mathsf{s}\backslash\mathsf{np})/\mathsf{np} : \lambda x.\lambda y.\mathbf{defeat}(y,x)$

The above two rules for functional application in fact indicate how the semantics is supposed to be assembled, viz., by *functional application*:

$(>\mathbf{A})\quad \mathsf{X}/\mathsf{Y} : f\ \ \mathsf{Y} : a \ \Rightarrow\ \mathsf{X} : fa$
$(<\mathbf{A})\quad \mathsf{Y} : a\ \ \mathsf{X}\backslash\mathsf{Y} : f \ \Rightarrow\ \mathsf{X} : fa$

f in the above two rules actually abbreviates $\lambda x.fx$, so that the resulting phrase on the right-hand side is in fact fa as a result of applying β-reduction to $(\lambda x.fx)(a)$.

Given these rule schemes, we can easily find a derivation for sentences, such as *Brazil defeats Germany*:

$$\begin{array}{ccc} \mathsf{np}{:}\mathbf{Brazil} & (\mathsf{s}\backslash\mathsf{np})/\mathsf{np}{:}\lambda x.\lambda y.\mathbf{defeat}(y,x) & \mathsf{np}{:}\mathbf{Germany} \\ \cline{2-3} \mathsf{np}{:}\mathbf{Brazil} & \multicolumn{2}{c}{\mathsf{s}\backslash\mathsf{np}{:}\lambda y.\mathbf{defeat}(y,\mathbf{Germany})} \\ \hline \multicolumn{3}{c}{\mathsf{s}{:}\mathbf{defeat}(\mathbf{Brazil},\mathbf{Germany})} \end{array}$$

A lot of linguistic phenomena can be perfectly handled by the two application rules. However, many researchers have argued that the AB calculus should be extended by rules that have a greater combinatory potential. CCG, for instance, employs rules for forward/backward (harmonic & crossed) composition, substitution, and type raising (we only list the forward versions):

Harmonic Composition $(>\mathbf{B})\quad \mathsf{X}/\mathsf{Y}\ \ \mathsf{Y}/\mathsf{Z} \ \Rightarrow\ \mathsf{X}/\mathsf{Z}$

Crossed Composition $(>\mathbf{B}_\times)\quad \mathsf{X}/\mathsf{Y}\ \ \mathsf{Y}\backslash\mathsf{Z} \ \Rightarrow\ \mathsf{X}\backslash\mathsf{Z}$

Substitution $(>\mathbf{S})\quad (\mathsf{X}/\mathsf{Y})/\mathsf{Z}\ \ \mathsf{Y}/\mathsf{Z} \ \Rightarrow\ \mathsf{X}/\mathsf{Z}$

Type Raising $(>\mathbf{T})\quad \mathsf{X} \ \Rightarrow\ \mathsf{Y}/(\mathsf{Y}\backslash\mathsf{X})$

Related to these rules are the three combinators (e.g., higher-order functions) for *composition* **B**, *subsitution* **S**, and *type raising* **T** (see (Steedman 2000)):

- $\mathbf{B}fg \equiv \lambda x.f(gx)$
- $\mathbf{S}fg \equiv \lambda x.fx(gx)$
- $\mathbf{T}x \equiv \lambda f.fx$

In a certain sense, even functional application can be seen as a combinator, since argument a can be regarded as a nullary function:

- $\mathbf{A}fa \equiv \lambda x.fx(a)$

The three combinators above indicate how semantics should be assembled within the categorial rules. Semantics construction is addressed later when we move to the TFS representation of the CCG rules.

Idea

The TFS encoding below distinguishes between atomic and complex categories. Atomic categories such as s do not have an internal structure. However, atomic categories in CCG are usually part of a structured inheritance lexicon, quite similar to HPSG. Atomic categories here do have a flat internal structure, encoding morpho-syntactical feature-value combinations. Thus, atomic categories in our transformation will be realized as typed feature structures to fully exploit the potential of typed unification.

In contrast, the most general functor category type has two subtypes / (*slash*) and \ (*backslash*) and defines three appropriate features: 1ST (FIRST), 2ND (SECOND), and MODE (for modalities, explained later). This encoding is similar to the CUG encoding in (Karttunen 1986; Uszkoreit 1986); however, the DIR (direction) feature is realized as a type, and the ARG (argument) and VAL (value) features through features 1ST and 2ND. Our encoding is advantageous in that it (i) makes a complex functor hierarchy possible, even multi-dimensional functors; (ii) allows for functors of more than two arguments, thus going beyond the potential of binary rules; and (iii) need not look at the directionality of the functor in order to specify the proper values for ARG and VAL (as is the case in Lambek's notation).

Underspecified atomic categories in the CCG rules above are realized through logic variables (coreferences) in the TFS rules below. Moreover, a distinguished list-valued feature DTRS (daughters) is employed in the TFS representation to model the LHS arguments of CCG rules.

Examples

We start with the TFS encoding of a proper noun, a transitive verb, and a modal verb, followed by the basic representation of the forward versions of the CCG rules, including a form of Lambda semantics in order to indicate how the compositional semantic approach of categorial grammars translates into a TFS grammar.

Lexicon Entries

A proper noun, such as *Germany* $\vdash \mathsf{np}$: **Germany** is mapped to a flat feature structure with distinguished attributes CAT and SEM:

$$\begin{bmatrix} \textit{germany} \\ \text{CAT } \textit{np} \\ \text{SEM } \mathbf{Germany} \end{bmatrix}$$

Actually, **Germany** is represented as a nullary function (i.e., a function with zero arguments), but this does not matter here. The value of SEM is either a function specification (type f) with NAME and ARGS features, or the representation of a Lambda term (type λ), encoded through VAR and BODY. The body of a Lambda term might again be a Lambda term or a function specification. Functional composition is encoded through an embedding of function specifications.

The representation of transitive verbs is a straightforward translation of the one-dimensional CCG specification *defeat* $\vdash \mathsf{(s\backslash np)/np} : \lambda x.\lambda y.\textbf{defeat}(y, x)$. Note that the de-curried representation suggests that β-reduction for x happens *before* y. Note further that even though x is bound first, it is the second argument of **defeat**.

$$\begin{bmatrix} \textit{defeat} & \\ \text{CAT} & \begin{bmatrix} / & \\ \text{1ST} & \begin{bmatrix} \backslash & \\ \text{1ST} & s \\ \text{2ND} & np \end{bmatrix} \\ \text{2ND} & np \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \lambda & \\ \text{VAR} & \boxed{x} \\ \text{BODY} & \begin{bmatrix} \lambda & \\ \text{VAR} & \boxed{y} \\ \text{BODY} & \begin{bmatrix} f & \\ \text{NAME} & \textbf{defeat} \\ \text{ARGS} & \left\langle \boxed{y}, \boxed{x} \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The representation of modal verbs is more complicated because P in the complex Lambda term below is not an argument like x (or x and y above), but instead a function that is *applied* to x—it might even be a Lambda term as the example *Brazil should defeat Germany* shows. Here is the categorial representation, followed by the TFS encoding: *should* $\vdash \mathsf{(s\backslash np)/(s\backslash np)} : \lambda P.\lambda x.\textbf{should}(Px)$

$$\begin{bmatrix} \textit{should} & \\ \text{CAT} & \begin{bmatrix} / & \\ \text{1ST} & \begin{bmatrix} \backslash & \\ \text{1ST} & s \\ \text{2ND} & np \end{bmatrix} \\ \text{2ND} & \begin{bmatrix} \backslash & \\ \text{1ST} & s \\ \text{2ND} & np \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \lambda & \\ \text{VAR} & \begin{bmatrix} \lambda & \\ \text{VAR} & \boxed{x} \\ \text{BODY} & \boxed{b} \end{bmatrix} \\ \text{BODY} & \begin{bmatrix} \lambda & \\ \text{VAR} & \boxed{x} \\ \text{BODY} & \begin{bmatrix} f & \\ \text{NAME} & \textbf{should} \\ \text{ARGS} & \left\langle \boxed{b}\ ] \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Rules

Next comes the rule for **Forward Functional Application**:

$(>\mathbf{A})\quad \mathsf{X/Y} : f \quad \mathsf{Y} : a \ \Rightarrow\ \mathsf{X} : fa$

$$\begin{bmatrix} >\mathbf{A} & \\ \text{CAT} & \boxed{X} \\ \text{SEM} & \boxed{f} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{CAT} & \begin{bmatrix} / & \\ \text{1ST} & \boxed{X} \\ \text{2ND} & \boxed{Y} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \lambda & \\ \text{VAR} & \boxed{a} \\ \text{BODY} & \boxed{f} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{CAT} & \boxed{Y} \\ \text{SEM} & \boxed{a} \end{bmatrix} \right\rangle \end{bmatrix}$$

Given this rule and the entries for *should*, *defeat*, and *Germany*, the twofold application of $(>\mathbf{A})$ yields the correct semantics for the VP *should defeat Germany*, viz., $\lambda x.\textbf{should}(\textbf{defeat}(x, \textbf{Germany}))$, or as a TFS, constructed via unification:

$$\begin{bmatrix} \lambda & \\ \text{VAR} & \boxed{x} \\ \text{BODY} & \begin{bmatrix} f & \\ \text{NAME} & \textbf{should} \\ \text{ARGS} & \left\langle \begin{bmatrix} f & \\ \text{NAME} & \textbf{defeat} \\ \text{ARGS} & \langle \boxed{x}, \textbf{Germany} \rangle \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

The TFS representation of the three rules to follow next are **Forward Harmonic Composition**, **Forward Substitution**, and **Forward Type Raising**. The motivation for such kind of rules, can, e.g., be found in (Baldridge 2002).

$(>\mathbf{B})\quad \mathsf{X/Y} : f \quad \mathsf{Y/Z} : g \ \Rightarrow\ \mathsf{X/Z} : \lambda x.f(gx)$

$$\begin{bmatrix} >\mathbf{B} & \\ \text{CAT} & \begin{bmatrix} / & \\ \text{1ST} & \boxed{X} \\ \text{2ND} & \boxed{Z} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \lambda & \\ \text{VAR} & \boxed{x} \\ \text{BODY} & \boxed{f} \begin{bmatrix} \text{ARGS}|\text{FIRST} & \boxed{g} \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{CAT} & \begin{bmatrix} / & \\ \text{1ST} & \boxed{X} \\ \text{2ND} & \boxed{Y} \end{bmatrix} \\ \text{SEM}|\text{BODY} & \boxed{f} \end{bmatrix}, \begin{bmatrix} \text{CAT} & \begin{bmatrix} / & \\ \text{1ST} & \boxed{Y} \\ \text{2ND} & \boxed{Z} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{VAR} & \boxed{x} \\ \text{BODY} & \boxed{g} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

$(>\mathbf{S})\quad \mathsf{(X/Y)/Z} : f \quad \mathsf{Y/Z} : g \ \Rightarrow\ \mathsf{X/Z} : \lambda x.fx(gx)$

$$\begin{bmatrix} >\mathbf{S} & \\ \text{CAT} & \begin{bmatrix} / & \\ \text{1ST} & \boxed{X} \\ \text{2ND} & \boxed{Z} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \lambda & \\ \text{VAR} & \boxed{x} \\ \text{BODY} & \boxed{f} \begin{bmatrix} \text{ARGS}|\text{REST}|\text{FIRST} & \boxed{g} \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{CAT} & \begin{bmatrix} / & \\ \text{1ST} & \begin{bmatrix} / & \\ \text{1ST} & \boxed{X} \\ \text{2ND} & \boxed{Y} \end{bmatrix} \\ \text{2ND} & \boxed{Z} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \lambda & \\ \text{VAR} & \boxed{x} \\ \text{BODY} & \boxed{f} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{CAT} & \begin{bmatrix} / & \\ \text{1ST} & \boxed{Y} \\ \text{2ND} & \boxed{Z} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \lambda & \\ \text{VAR} & \boxed{x} \\ \text{BODY} & \boxed{g} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

$(>\mathbf{T})\quad \mathsf{X} : x \ \Rightarrow\ \mathsf{Y/(Y\backslash X)} : \lambda f.fx$

$$\begin{bmatrix} \mathbf{>T} & \\ \textsc{cat} & \begin{bmatrix} / & \\ \textsc{1st} & \boxed{\mathsf{Y}} \\ \textsc{2nd} & \begin{bmatrix} \backslash & \\ \textsc{1st} & \boxed{\mathsf{Y}} \\ \textsc{2nd} & \boxed{\mathsf{X}} \end{bmatrix} \end{bmatrix} \\ \textsc{sem} & \begin{bmatrix} \lambda & \\ \textsc{var} & \boxed{\mathsf{f}} \\ \textsc{body} & \begin{bmatrix} f & \\ \textsc{name} & \boxed{\mathsf{f}} \\ \textsc{args} & \langle \boxed{\mathsf{x}} \rangle \end{bmatrix} \end{bmatrix} \\ \textsc{dtrs} & \left\langle \begin{bmatrix} \textsc{cat} & \boxed{\mathsf{X}} \\ \textsc{sem} & \boxed{\mathsf{x}} \end{bmatrix} \right\rangle \end{bmatrix}$$

Extensions

In this section, we outline several extensions of the basic CG system and show what their TFSs representation look like.

$-Convention & Generalized Forward Composition

The VP *should defeat Germany* from the rule section can not only be analyzed by a twofold application of ($>\mathbf{A}$), but also by applying ($>\mathbf{B}$) to *should* and *defeat*, followed by ($>\mathbf{A}$). Now, ($>\mathbf{B}$) must be generalized in case we are even interested in ditransitive verbs, or even VPs with further PP attachments. Instead of describing every possible alternative, (Steedman 2000) devised a compact notation using $-schemes to characterize functions of varying numbers of arguments, or as (Baldridge 2002) puts it: "In essence, the $ acts as a stack of arguments that allows the rule to eat into a category". For example, the schema $\mathsf{s}/\$$ is a representative for the infinite set $\{\mathsf{s}, \mathsf{s}/\mathsf{np}, (\mathsf{s}/\mathsf{np})/\mathsf{np}, \ldots\}$.

Formally, the expansion of a $-category can be inductively defined as follows. Let C be the set of complex categories, as defined earlier, F_2 the set of binary functor symbols, and let $c \in \mathsf{C}$ and $\mathsf{f} \in \mathsf{F}_2$. Define $\mathsf{C}_\epsilon := C \cup \{\epsilon\}$, $\mathsf{cf}\epsilon := \mathsf{c}$, and $\mathsf{cfC}_\epsilon := \{\mathsf{cfd} \mid \mathsf{d} \in \mathsf{C}_\epsilon\}$. Then $\mathsf{cf}\$:= (\mathsf{cfC}_\epsilon)\mathsf{fC}_\epsilon$.

Let us move on to the rule for generalized forward composition ($>\mathbf{B}^n$) which employs $ and its TFS counterpart:

$$(>\mathbf{B}^n) \quad \mathsf{X}/\mathsf{Y} \quad (\mathsf{Y}/\mathsf{Z})/\$ \;\Rightarrow\; (\mathsf{X}/\mathsf{Z})/\$$$

$$\begin{bmatrix} >\mathbf{B}^{n>1} & \\ \textsc{cat} & \begin{bmatrix} / & \\ \textsc{1st}^{n-1} & \begin{bmatrix} / & \\ \textsc{1st} & \boxed{\mathsf{X}} \\ \textsc{2nd} & \boxed{\mathsf{Z}} \end{bmatrix} \\ \textsc{2nd} & \boxed{\$} \end{bmatrix} \\ \textsc{dtrs} & \left\langle \begin{bmatrix} \textsc{cat} & \begin{bmatrix} / & \\ \textsc{1st} & \boxed{\mathsf{X}} \\ \textsc{2nd} & \boxed{\mathsf{Y}} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \textsc{cat} & \begin{bmatrix} / & \\ \textsc{1st}^{n-1} & \begin{bmatrix} / & \\ \textsc{1st} & \boxed{\mathsf{Y}} \\ \textsc{2nd} & \boxed{\mathsf{Z}} \end{bmatrix} \\ \textsc{2nd} & \boxed{\$} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

The above TFS uses a "coordinated" path expression $\textsc{1st}^{n-1}$ at two places inside the rule structure and is, in a certain sense, even worse than *functional uncertainty* (Kaplan & Maxwell III 1988), since it involves counting. To the best of our knowledge, we are not aware of TFS formalisms which offer such descriptive means. We thus understand the above structure as a schema that can be compiled into $k - 1$ different concrete rules for $1 < n \leq k$. Another way to carry over the meaning would be to add a unary and a binary helper rule for each $-rule which together simulate the expansion of a $-category. We have opted for the first solution, since the latter could blow up the search space of the parser.

We finally note that $>\mathbf{B}^1$ is equivalent to the original rule $>\mathbf{B}$. In case we define $\textsc{1st}^0 := \epsilon$ and assume that $\textsc{2nd} \doteq \boxed{\mathsf{Z}} \wedge \textsc{2nd} \doteq \boxed{\$}$ leads to $\boxed{\mathsf{Z}} = \boxed{\$}$ (features are functional relations!), there is no need to specify $>\mathbf{B}^1$ separately.

In principle, other rule schemata might be generalized in such a way, but at the expense of further uncertainty and overgeneration during parsing.

Atomic Categories & Morpho-Syntax

As indicated earlier, atomic categories in CCG usually do have a flat internal structure. For instance, the category s_i refers to an inflection phrase (Baldridge 2002). The TFS representation then uses s_i as a type, having the following definition:

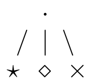
$$\mathsf{IP} \equiv \begin{bmatrix} s_i & \\ \textsc{spec} & \textit{boolean} \\ \textsc{ant} & \textit{boolean} \\ \textsc{case} & \textit{case} \\ \textsc{vform} & \textit{fin} \\ \textsc{marking} & \textit{unmarked} \end{bmatrix}$$

Words in CCG usually refer to these more specialized categories; for instance, the ECM verb *believe* $\vdash (\mathsf{s}_i \backslash \mathsf{np})/\mathsf{s}_{fin}$. Given such specific category information, TFS unification takes care that the additional constraints are "transported" throughout the derivation tree.

Modes & Modalized CCG

Besides having more control through specialized atomic categories as shown above, *multi-modal CCG* incorporates means from *Categorial Type Logic* to provide a further fine-grained lexical control through so-called *modalities*; see (Baldridge & Kruijff 2003) for a detailed description. For example, the complex category of the coordination particle *and* $\vdash (\mathsf{s}_i \backslash \mathsf{s}_i)/\mathsf{s}_i$ which can lead to unwanted analyses is replaced by the modalized category $(\mathsf{s}_i \backslash_\star \mathsf{s}_i)/_\star \mathsf{s}_i$.

In principle, modes can be "folded" into subtypes of the very general complex category types / and \. We have, however, opted for an additional feature MODE which takes values from the following atomic mode type hierarchy:



There are further modalities which are not of interest to us here. Let us finally present the TFSs for *and* and the multi-modal CCG forward type raising rule rule ($>\mathbf{T}$) which even enforces modes to be identical between the embedded and the outer slash.

$$\begin{bmatrix} \textit{and} & \\ \textsc{cat} & \begin{bmatrix} / & \\ \textsc{1st} & \begin{bmatrix} \backslash & \\ \textsc{1st} & s_i \\ \textsc{2nd} & s_i \\ \textsc{mode} & \star \end{bmatrix} \\ \textsc{2nd} & s_i \\ \textsc{mode} & \star \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} >\mathbf{T} & \\ \textsc{cat} & \begin{bmatrix} / & \\ \textsc{mode} & \boxed{M} \\ \textsc{1st} & \boxed{Y} \\ \textsc{2nd} & \begin{bmatrix} \backslash & \\ \textsc{mode} & \boxed{M} \\ \textsc{1st} & \boxed{Y} \\ \textsc{2nd} & \boxed{X} \end{bmatrix} \end{bmatrix} \\ \textsc{dtrs} & \langle [\, \textsc{cat}\ \boxed{X}\,] \rangle \end{bmatrix}$$

First Measurements

We have compared the performance of the CCG parser and the PET system on a MacBook Pro (2GHz Core Duo, 32 bit architecture). The measurements were carried out against a hand-crafted artificial test corpus of 5,000 sentences with an average length of 7 and a maximal length of 12 words, including sentences with heavy use of different kinds of coordination, such as *Brazil will meet and defeat Germany* or *Brazil should defeat Germany and Italy and England.*

We have switched off the semantics and have only compared the syntactic coverage, using categorial information, including modes. We have also switched off the type raising rules in both parsers, since the OpenCCG parser seems to ignore them in analyses licensed by the grammar theory. Packing in both parsers has been switched on, supertagging switched off (in fact, PET does not provide a supertagging stage).

We further note that we have obtained about twice as much analyses for PET (approximately 15,000 analyses) as the OpenCCG system, the reason for this currently unclear. For instance, the CCG parser produces only **one** analysis for the sentence *Brazil should defeat Germany*, even though a careful inspection of the rules shows that **two** analyses are possible (as is the case for PET), viz.,

$$[(<\mathbf{A})\ \textit{Brazil}\ [(>\mathbf{A})\ \textit{should}\ [(>\mathbf{A})\ \textit{defeat}\ \ \textit{Germany}\,]\,]\,]$$

$$[(<\mathbf{A})\ \textit{Brazil}\ [(>\mathbf{A})\ [(>\mathbf{B})\ \textit{should}\ \ \textit{defeat}\,]\ \textit{Germany}\,]\,]$$

Even though we double the number of analyses, PET is about one magnitude faster (overall 2.67 vs. 28.9 seconds for the full set of 5,000 sentences).

Both PET and the OpenCCG system have implemented standard CYK parsers. We believe that the difference in the running time is related to the choice of the programming language (C++ vs. Java), but also to maintenance effort and the still ongoing development of the PET system by an active community, whereas the evolution of the core parsing engine in the OpenCCG library seems to have ended several years ago.

To some extend, the above mismatch is related to the fact that certain "settings" in the CCG are realized through *program code*, but **not** *declaratively* stated in the lingware. For instance, the type raising rules can in principle be applied to arbitrary categories, but, by default, the OpenCCG code limits them to NPs only. Given our treatment, such a restriction can be easily stated in the TFSs for the type raising rules, and we think that this is the right place to do so:

$$\begin{bmatrix} >\mathbf{T} & \\ \textsc{cat} & \begin{bmatrix} / & \\ \textsc{1st} & \boxed{Y} \\ \textsc{2nd} & \begin{bmatrix} \backslash & \\ \textsc{1st} & \boxed{Y} \\ \textsc{2nd} & \boxed{X} \end{bmatrix} \end{bmatrix} \\ \textsc{dtrs} & \langle [\, \textsc{cat}\ \boxed{X}\, \textit{np}\,] \rangle \end{bmatrix}$$

Other "adjusting screws" in OpenCCG, e.g., the specification of the atomic mode hierarchy (see last subsection) are also "casted" in program code (deeply nested *if-then-else* statements), whereas our treatment uses a type hierarchy, helping to better understand and manipulate the parser's output. Given these remarks, explaining missing analyses in OpenCCG has required a deep inspection of the program code. Besides the MODE dimension, we found a further orthogonal binary ABILITY dimension with values *inert* and *active* that was hidden in the program code (Java classes) for each categorial rule. The PET version of CCG still overgenerates (to a lesser extent) and we hope to unveil the secrets at the conference.

Moving Further

The transformation schema described in this paper has been manually constructed for the rules, the lexical types, and a small set of lexicon entires. In order to automatically transform the OpenCCG grammars from our Lab for English and Italian, we have implemented code that operates on the XML output of the `ccg2xml` converter for CCG's WebCCG input format. This includes files for rules, general types, and so-called families which are collections of lexical types and corresponding lexical entries.

Contrary to traditional CG and CCG, OpenCCG does not use Lambda semantics, but instead comes with a kind of Davidsonean event semantics, comparable to MRS, building on Blackburn's hybrid modal logic: Hybrid Logic Dependency Semantics (HLDS) (Baldridge & Kruijff 2002). Looking more closely on the seemingly different notation, it becomes quite clear that HLDS formulae can be straightforwardly translated into a TFS representation. We can only throw a glance on a small example at the end of this paper.

Originally, the HLDS representations were built up in tandem with the construction of the categorial backbone (Baldridge & Kruijff 2002), comparable to the construction of Lambda semantics in our rules before. (White & Baldridge 2003) has improved on this construction by attaching the semantics, i.e., the elementary predications (EPs), directly to the atomic categories from which a complex category is built up (see (Zeevat 1988) for a similar treatment in UCG).

Consider the sentence *Marcel proved completeness* (Kruijff & Baldridge 2004). Subscripts attached to atomic categories (the nominals) can be used to access them. The satisfaction operator @ that is equipped with a subscript e indicates that the formulae to follow hold at a state named e:

$$\textit{proved} \vdash (\mathsf{s}_e \backslash \mathsf{np}_x)/\mathsf{np}_y : \\ \quad @_e\textbf{prove} \wedge @_e\langle\text{TENSE}\rangle\textbf{past} \wedge @_e\langle\text{ACT}\rangle x \wedge @_e\langle\text{PAT}\rangle y$$

$$\textit{Marcel} \vdash \mathsf{np}_m : @_m\textbf{Marcel}$$

$$\textit{completeness} \vdash \mathsf{np}_c : @_c\textbf{completeness}$$

By *conjoining* the EPs during the application of $(>\mathbf{A})$ and $(<\mathbf{A})$, we immediately obtain

$$\textit{Marcel proved completeness} \vdash \mathsf{s}_e : \\ \quad @_e\textbf{prove} \wedge @_e\langle\text{TENSE}\rangle\textbf{past} \wedge @_e\langle\text{ACT}\rangle m \wedge \\ \quad @_e\langle\text{PAT}\rangle c \wedge @_m\textbf{Marcel} \wedge @_c\textbf{completeness}$$

Exactly these effects can be achieved through unification in our framework. The CCG nominals are realized through logic variables (coreference tags), atomic categories, such as s or np are assigned a further feature INDEX, cospecified with the semantics, and the nominals are realized through ordinary features. In theory, SEM is a set-valued feature whose elements are combined conjunctively (as in HLDS or MRS). Since $\mathcal{TDL}$ (and PET) does not provide sets, the usual list implementation is used. This gives us the following TFSs (we have omitted the explicit representation of the name of the event variables e, m, and c in the individual EPs below):

$$\begin{bmatrix} \textit{proved} & \\ \text{CAT} & \begin{bmatrix} / & \\ \text{1ST} & \begin{bmatrix} \backslash & \\ \text{1ST} & \begin{bmatrix} \textit{s} \\ \text{INDEX}\ \boxed{e} \end{bmatrix} \\ \text{2ND} & \begin{bmatrix} \textit{np} \\ \text{INDEX}\ \boxed{x} \end{bmatrix} \end{bmatrix} \\ \text{2ND} & \begin{bmatrix} \textit{np} \\ \text{INDEX}\ \boxed{y} \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \boxed{e}\ \langle \textbf{prove}, [\text{TENSE}\ \textit{past}], [\text{ACT}\ \boxed{x}], [\text{PAT}\ \boxed{y}] \rangle \end{bmatrix}$$

$$\begin{bmatrix} \textit{marcel} & \\ \text{CAT} & \begin{bmatrix} \textit{np} \\ \text{INDEX}\ \boxed{m} \end{bmatrix} \\ \text{SEM} & \boxed{m}\ \langle \textbf{Marcel} \rangle \end{bmatrix} \quad \begin{bmatrix} \textit{completeness} & \\ \text{CAT} & \begin{bmatrix} \textit{np} \\ \text{INDEX}\ \boxed{c} \end{bmatrix} \\ \text{SEM} & \boxed{c}\ \langle \textbf{completeness} \rangle \end{bmatrix}$$

Alternatively, the list representation of EPs might be replaced by a single complex feature structure. However, the list implementation makes it easy to implement relational information, e.g., the representation of several modifiers. Given the above encoding, there is no longer a need to specify semantics construction in each of the categorial rule schemata: semantics construction simply "happens" here when categorial information is unified. In a certain sense, this is easier and more elegant than representing the effects of the different combinators **A**, **B**, **S**, **T** in the different kinds of rule schemata, as we have described in the beginning of this paper. More complex constructions involving, e.g., coordination particles, stipulate that the list under SEM is in fact a difference list in order to ease the implementation of a list append that is not required in the example above.

References

Baldridge, J., and Kruijff, G.-J. M. 2002. Coupling CCG and hybrid logic dependency semantics. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 319–326.

Baldridge, J., and Kruijff, G.-J. M. 2003. Multi-modal combinatory categorial grammar. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, 211–218.

Baldridge, J. 2002. *Lexically Specified Derivational Control in Combinatory Categorial Grammar*. Ph.D. Dissertation, University of Edinburgh, Division of Informatics, Institute for Communicating and Collaborative Systems.

Barendregt, H. 1984. *The Lambda Calculus, its Syntax and Semantics*. Amsterdam: North-Holland.

Bouma, G. 1988. Modifiers and specifiers in categorial unification grammar. *Linguistics* 26:21–46.

Calder, J.; Klein, E.; and Zeevat, H. 1988. Unification categorial grammar: A concise, extendable grammar for natural language processing. In *Proceedings of the 12th International Conference on Computational Linguistics*, 83–86.

Callmeier, U. 2000. PET—A Platform for Experimentation with Efficient HPSG Processing. *Natural Language Engineering* 6(1):99–107.

Kaplan, R. M., and Maxwell III, J. T. 1988. An algorithm for functional uncertainty. In *Proceedings of the 12th International Conference on Computational Linguistics*, 297–302.

Karttunen, L. 1986. Radical lexicalism. Technical Report CSLI-86-68, Center for the Study of Language and Information, Stanford University.

Krieger, H.-U. 1995. *$\mathcal{TDL}$—A Type Description Language for Constraint-Based Grammars. Foundations, Implementation, and Applications*. Ph.D. Dissertation, Universität des Saarlandes, Department of Computer Science.

Krieger, H.-U. 2007. From UBGs to CFGs—a practical corpus-driven approach. *Natural Language Engineering* 13(4):317–351. Published online in April 2006.

Kruijff, G.-J. M., and Baldridge, J. 2004. Generalizing dimensionality in combinatory categorial crammar. In *Proceedings of the 20th International Conference on Computational Linguistics*.

Steedman, M. 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

Uszkoreit, H. 1986. Categorial unification grammars. In *Proceedings of the 11th International Conference on Computational Linguistics*, 187–194.

White, M., and Baldridge, J. 2003. Adapting chart realization to CCG. In *Proceedings of the 9th European Workshop on Natural Language Generation*.

Zeevat, H. 1988. Combining categorial grammar and unification. In Reyle, U., and Rohrer, C., eds., *Natural Language Parsing and Linguistic Theories*. Reidel, Dordrecht. 202–229.