Theory and applications in Computational Linguistics - Is there common ground?

Wolfgang Menzel
Universität Hamburg

Abstract

Theoretical reasoning and application-oriented system development are frequently experienced as quite opposite strategic orientations within most areas of Computational Linguistics. Goals, methods and results differ to a degree which prevent them from being easily reconciled. Although the the current state seems to render any attempt to bridge this gap almost hopeless, this paper nevertheless adopts an optimistic position in favour of a convergence of the two positions. Drawing on recent developments in optimality theory and constraint-based reasoning, it illustrates how a paradigm shift away from the consistency-centered approaches of unification-based grammar towards an active treatment of inconsistency in language processing might open up new possibilities for more close-to-reality applications. There is a growing body of evidence supporting the claim that both perspectives are indispensable for further advances in our field: A strong theoretical foundation of application-oriented research *and* a theory which does not neglect application-specific concerns.

Nichts ist praktischer als eine gute Theorie. L. Boltzmann

1 Do we need linguistic theory for applications of CL?

Early developments in the area of Computational Linguistics have been strongly influenced by theories from Cognitive Linguistics and their underlying formal frameworks. The strong position of context-free models in syntax (Chomsky, 1957; Chomsky, 1965; Chomsky, 1981) and the dominance of logically oriented approaches to semantics and discourse (Montague, 1974; Kamp, 1981) can be taken as only two particular examples for this general phenomenon. Together with the theory, however, some important methodological assumptions have been inherited from linguistics as well, most notably the central role of the usually introspectively grounded notion of well-formedness. This obviously has lead to a strong bias towards the study of selected linguistic phenomena, focussing on a number of carefully constructed test cases (e.g. some peculiar movement phenomena or the scope ambiguity of quantification), although such phenomena hardly can be observed in real language use.

So it came to no surprise that eventually harsh criticism against such an unworldly kind of activities has been raised from a more application-oriented perspective (Nagao, 1988; Heyer, 1988; Heyer, 1989). As a consequence the call to strengthen research paradigms with a clear commitment to practically relevant tasks like machine translation, document retrieval or information extraction became much stronger. In particular, the criticism pointed out all those questions which might be theoretically interesting but are completely irrelevant for practical purposes, because of being either too expensive or too peripheral. On the other hand it tried to draw the attention of the research community to the abundance of very practical ("dirty") problems like abbreviations, punctuation or typos, which usually are considered less rewarding and uninteresting from a theoretical point of view.

In all these constributions it was noted repeatedly that coverage and robustness have to be treated as much more important issues than elegance and universality of a formal solution. Moreover, the very foundation of linguistic theory, the fundamental distinction between well-formed and deviant utterances, was questioned seriously for the first time: In contrast to an abstract grammar model, a natural language system is not designed "to accept or reject input strings", but instead should "assign the most reasonable structure to every input string, and to comment on it, when necessary." (Jensen in Nagao (1988)). Therefore all and not only the grammatical constructions of a natural language need to be considered. Although in general the contribution of linguistic theory has been acknowledged, a reorientation was strongly recommended: "We are not skeptical of theory but of existing theories" (Jensen in Nagao (1988)).

2 Do data suffice?

Since that time much has changed. The empiricist revolution promoted a variety of stochastic and connectionist techniques and apparently moved Computational Linguistics even farther away from the mainstream of Cognitive Linguistics. Here, a major driving force has been the availability of increasingly large corpora of written and spoken language together with a tremendous decrease in the cost of computational resources. Interestingly enough, this paradigm shift was accompagnied by a shift in research topics as well, which however was not so much guided by new insights into the problem but more by the power of the newly available technical means: Recursively defined structural descriptions have mainly been replaced by rather flat representations (e.g. syntactic or semantic tags) (Wermter and Weber, 1997), which are even used for purposes where structural relationships always have been considered crucial, like in Machine Translation (Brown et al., 1990). Only recently have some attempts towards more powerful means of representation been pursued (Chelba and Jelinek, 1998; Henderson and Lane, 1998).

Unfortunately, the empiricist revolution also did not lead to the sometimes already expected breakthrough (Clark and Lutz, 1990). Stochastic approaches to part-of-speech tagging are still lagging behind the more conventional symbolic solutions (Abney, 1997) and only in very specialised domains are trainable models able to achieve a close-to-competitive performance. On the other hand, corpus-based techniques offer a quite attractive solution to the knowledge acquisition bottleneck from which all symbolic approaches suffer. Moreover, due to their ability to extract generalised knowledge from a collection of data, such methods provide a certain kind of robustness against unseen and thus unexpected events.

Certainly, there is still a lot to be expected from the application of soft-computing techniques in Computational Linguistics, since this area of research still looks pretty far from having exhausted its potential already. Nevertheless, it becomes more and more obvious that these attempts alone do not provide the ultimate solution to all the pending problems of language processing. Using empiricist methods does not automatically result in better solutions, but usually breaking new grounds becomes possible. It first of all was this observation, which initiated attempts to combine the partially complementory advantages of symbolic and subsymbolic techniques in hybrid architectures to increase the overall performance (Wahlster, Bub, and Waibel, 1997).

Another important observation concerns a kind performance saturation which happens to appear with any kind of probabilistic or connectionist model: From a certain point onwards the quality of the computation results does not improve much further, irrespective of the amout of training data used. This effect is certainly not the same as the one which can be found for symbolic techniques, where an exponential increase of development effort is typical: Trying to achieve the last n percent of coverage is usually much more expensive than what has been spent so far. The problem with stochastic approaches, however, is a more fundamental one. On the one hand, the very nature of probabilistic models allows them to handle the most frequent cases fairly well. Therefore impressive inital performance figures can often be achieved with a comparatively low investment of effort. On the other hand, it follows from a fundamental law of information theory that most informational content is contained in the rare cases, which differ somehow from the vast majority of "normal" ones. Moreover, techniques for supervised training are still restricted to rather simple model structures, which make it inherently difficult to capture all the relevant distinctions.

Finally, the undeniable advantages of probabilistic and connectionist models with respect to robustness and trainability have been gained at the expense of two other, certainly not less important drawbacks: A strong dependence on the training data, which hardly allows a transfer of components to other application domains, and a dramatic loss of explanatory capacity, which strongly questions the scientific nature of such efforts.

Nevertheless, we owe at least one major insight to the empiricist revolution, which has an importance to the field in general that can hardly be overestimated. By demonstrating that numerical annotation and arbitration schemes can be effectively used to decide between competing interpretation alternatives, they introduced the new paradigm of plausibility-based reasoning into natural language processing and thus made a decisive contribution towards solutions with a higher degree of robustness.

3 Can the gap be bridged?

Being able to explain certain observable phenomena can be considered one of the primary concerns of theoretical reasoning. Trying to achieve this goal, it asks questions about the conditions under which a particular phenomenon occurs, and the causal relationships between external factors and the observable behaviour.

Unfortunately, corpus-based methods do not have much to contribute in this respect. In extreme cases, the quality of a model and its explanatory power are quite opposite to each other. Perhaps, this negative correlation can be observed most convincingly with corpus-based methods for speech synthesis (Stöber et al., 1999), which achieve a remarkable level of quality both in terms of intelligibility and naturalness, but obviously tell us little about the nature of speech itself. On the other hand, solutions based on very detailed knowledge about the structure of human speech signals suffer from a disappointingly low quality, which has excluded them almost completely from consideration for any practical application. Although this relationship is far from being as dramatic at other levels of language processing, the opacity of probabilistic and connectionist approaches makes them considerably less attractive from a scientific point of view.

Theoretical reasoning is a procedure of abstraction, where generalised propositions can only be found at the expense of supressing more and more detail, which can be considered as irrelevant at least for that specific purpose. On the other hand, software system development is essentially a process of abstraction as well. Therefore the crucial question in both cases relates to whether the abstractions that have been (often implicitly) adopted can really be considered acceptable.

In the ideal case any abstraction should be guided by clearly specified goals in mind. In reality, however, this process is equally much if not even much stronger influenced by the available means. This became obvious for the data-driven approaches. It holds, however, for the construction of a theory also, where the available knowledge about mathematical theories, thoroughly investigated formal structures and efficient computational paradigms usually plays a highly influential role.

This can be exemplified very well by the development of grammatical theories during the last five decades. Here, it was the availability of mathematical results on rewriting systems, first order logics, and mechanical deduction, which influenced the theoretical thinking in many areas of linguistic inquiry to a considerable degree. This school of reasoning brought about a number of models for grammar and meaning and eventually led to the development of the very powerful paradigma of unification-based formalisms (Shieber, 1992).

If linguistic data are represented as simple sequences (lists) of lexical items like

[john, arrived],

a coherent translation into a first-order interpretation can be given for concepts such as context-free rewriting rules and the lexicon, where grammar rules correspond to universally quantified formula and lexical items to simple propositions (cf. Table 1).

Both parsing and generation can then be conceived of as logical deduction. Due to its firm grounding in formal logics this development, however, passes on a major disadvantage

S	$\rightarrow$	NP VP		$\forall$ x y z . np(x) $\wedge$ vp(x) $\wedge$ append(x,y,z)	$\rightarrow$	s(z)
NP	$\rightarrow$	PN		$\forall$ x . pn(x)	$\rightarrow$	np(x)
VP	$\rightarrow$	Vintr		$\forall$ x . vintr(x)	$\rightarrow$	vp(x)
Vintr	$\rightarrow$	arrived		vintr([arrived])		
PN	$\rightarrow$	john		pn([john])		

Table 1: Context-free grammar rules and their first-order equivalences.

to all its descendants, which makes them so vulnerable in almost any practical setting: It is the central notion of consistency all thinking in classical logics is based upon. With its inability to derive any meaningful conclusion from an inconsistent logical theory it puts a very strong requirement on every kind of domain to be modelled, since it needs to be perceived as also being consistent. Unfortunately, natural language does not fit this assumption quite well.

Enforcing a consistency requirement on a formal model of natural language amounts to ensuring the union of all axioms of the grammar $\mathcal{G}$, the lexicon $\mathcal{D}$ and all possible utterances s of a language $\mathcal{L}$ being completely contradiction-free:

$$\exists \mathcal{G} \, \exists \mathcal{D} \, \forall s \in \mathcal{L} \ . \ \mathcal{G} \cup \mathcal{D} \cup \{s\} \not\models \bot$$

As one can easily find out, even for highly idealised notions of language general consistency cannot easily be established. Quite to the contrary, inconsistencies are pervasive in natural language. They occur within and between almost every descriptive level. Against this background the fundamental competence-performance distinction looks more like being a construct which has been invented to make the data fit the available formal means, rather than providing a helpful device to model real human language capabilities.

In syntax, for instance, contradictions can be found e.g. for

- valence requirements, e.g. the optionality of the direct object, which is typical for many transitive verbs,
- ordering regularities, e.g. the position of the finite verb as being either sentence initial, second or final, or
- the extraposition of relative clauses, where proximity requirements of the verb group strongly interfere with the desire to minimise referential distance.

Furthermore, constituent ordering in the German *mittelfeld* frequently shows a conflict between the unmarked precedence (subject initial) and special emphatic needs (topicalisation).

Conflicting evidence plays an even more important role on the semantic level. Selectional restrictions, for instance, are especially conflict-prone, since semantic types are often subject to processes like type-shift and metaphorical reinterpretation. Even stronger might be the tension between the propositional content of an utterance and the available background knowledge, where any consistency requirement would render the communication of novel content infeasible.

Now, it eventually becomes obvious that inconsistencies are far from being an undesirable deviation from the core of natural language proper. They rather constitute one of the fundamental mechanisms interpersonal communication heavily relies upon. Speakers use techniques like topicalisation and metaphorics systematically to make their contributions more interesting and more attractive to the attention of the listener. Such a behaviour should come to no surprise, as we already noticed that according to the very definition of entropy the usually less frequent deviation from the normal case carries much more informational content than the normal case itself. The ability of human listeners to notice such deviations also brings them in a position to understand an unfinished sentence with e.g. unsatisfied open obligatory requirements as an invitation to take the initiative in a face-to-face conversation. And finally, inconsistencies are systematically used to produce language-based humour by highlighting a difference between a contextually highly plausible interpretation and another one, being less likely and somewhat strange in the given context.

Other examples of contradictory regularities (which obviously come along with a less direct communicative relevance) can be found in the area of pronominal reference, where the necessity to ensure a minimal referential distance if often in conflict with other preferences like syntactic or semantic parallelism. In some, albeit rare cases the agreement between pronoun and antecedent might even switch from grammatical to natural gender, as e.g. in:

Das Mädchen$_i$ sucht ihre$_i$ Puppe.

The most troublesome aspect of the general consistency requirement certainly is the necessity to extend it to the usually open ended set of possible utterances, which excludes ungrammatical, nonsense or other "deviant" input, as well as any form of language change from consideration. In all these cases the linguistic data is in conflict with the language system and without special provision no interpretation can be obtained for the corresponding utterance.

Unfortunately, being restricted to strictly well-formed input is the fundamental assumption most linguistic theories are based upon. With their strong, competence-based focus on idealised language they abstract away such input-related inconsistencies and are by no means able to explain the processing of deviant utterances. Apart from any consequences for building computational applications this situation should give rise to doubts about the cognitive adequacy of such an approach as well. Since human listeners are capable to analyse slightly ill-formed utterances without any apparent difficulty (in many cases even without noticing an error at all) the question needs to be asked, whether a model which is unable to reflect such a fundamental capability at least partially can be considered a good model of the human language comprehension system.

Inconsistencies can be introduced into a logical theory by using various non-monotonic reasoning techniques instead of the traditional monotonic ones. One of the most simple approaches is default reasoning which tries to retract specific assumptions about the normal behaviour of a certain model component in case a conflict was detected. Retraction can be applied

- to certain constructions in the grammar (constraint retraction) e.g. dropping an agreement requirement if it cannot be satisfied (Douglas and Dale, 1992),
- to certain properties of the input data (e.g. partial parsing (Lavie, 1996) or
- to both (Mellish, 1989).

In fact, default reasoning does not attempt to include inconsistencies into the grammar but tries to determine the conditions (usually called extensions of the theory) under which consistency can be reestablished. If several competing possibilities are available, only those which require a minimal set of retractions are considered. The difficulty with this kind of approach is that the number of possibilities grows exponentially with the number of defeasibly propositions and there is little information available to guide the search for promising solutions. This deficiency, however, might be overcome by means of soft-computing techniques, which extend the binary decision about a proposition being retractable or not to arbitrarily fine grained gradings. Now, it becomes possible to arbitrate evidence of different strength against each other and to select an optimal theorem (e.g. a parse) given the axioms for the utterance, the lexicon and the grammar. To apply the idea to the traditional notion of a rule-based grammar, however, introduces additional difficulties.

- The local domain to a context-free rule is not large enough to model the necessary generalisations with a sufficiently high level of reliability (Bod, 1995).
- To express the different pieces of preferential knowledge requires a sufficiently rich representation, where the combination of several preferences leads to an exponential increase in the number of rules.

The second point can easily be illustrated by a rather simplistic example in which we consider a binary branching structure of the type

```
  A
 / \
B   C
```

where the ordering of the daughter nodes should be free but subject to a certain preference. To model this situation we need already two context-free rules, namely

$$\begin{array}{llll} p_1: & A & \rightarrow & B\ C \\ p_2: & A & \rightarrow & C\ B \end{array}$$

where p_1 and p_2 denote a kind of preference factor or probability. If now the precedence of the two daughter nodes depends additionally on a binary feature available at node B the number of rules required doubles:

$$\begin{array}{llll} p_{11}: & A & \rightarrow & B_1\ C \\ p_{22}: & A & \rightarrow & B_2\ C \\ p_{21}: & A & \rightarrow & C\ B_1 \\ p_{22}: & A & \rightarrow & C\ B_2 \end{array}$$

This duplication becomes necessary only to be able to attach the new preference factors although they might also be computed by combining those from the underlying ordering constraints. If another such feature on either B or C is available the number of rules doubles once again. Note that exactly this kind of preferential constraints seems to rule the German mittelfeld, where the acceptability of a specific sequentialisation is (among other factors) influenced by the filler of a valency being pronominalised or not. Let us assume that a pairwise comparison of the four topicalised and non-topicalised versions of the subordinate clauses (1a) through (4b)

(1a) ... weil die Mutter dem Kind vertraut.
(1b) ... weil dem Kind die Mutter vertraut.
(2a) ... weil sie dem Kind vertraut.
(2b) ... weil dem Kind sie vertraut.
(3a) ... weil die Mutter ihm vertraut.
(3b) ... weil ihm die Mutter vertraut.
(4a) ... weil sie ihm vertraut.
(4b) ... weil ihm sie vertraut.

yield roughly the following acceptability judgements

$$(1a) > (1b),\ (2a) \gg (2b),\ (3a) \sim (3b) \text{ and } (4a) \gg (4b)$$

where $\sim$ denotes comparable, $>$ better and $\gg$ much better acceptability.

Instead of providing eight rules to cover all ordering preferences, a similar result can be obtained by means of three partially inconsistent constraints

p_3: the subject must precede the object
p_4: only material in topic position should be pronominalised
p_5: a pronominalised subject must precede a pronominalised object

which for equally sized preference factors p_i produce differences in scoring for the four sentence pairs of one, two, zero, and two respectively, that mirror the intuitive acceptability judgements given above surprisingly well.

What underlies this kind of grammar modelling is an attempt to decouple the conditions about the well-formedness of an utterance from the space of possible structural hypotheses, two sides of a coin which are inseparably glued together in a context-free rule: Only sentences generated by rules of the grammar are well-formed, and only well-formed sentences are generated by rules. To separate these two aspects is not a particularly new idea. It has already been achieved in constraint-based, but binary-valued models of grammar (Shieber, 1992), which however remained strongly committed to the consistency notion of first order logic. One of the first attempts to escape from this consistency cage has been tried within the framework of optimality theory (Prince and Smolensky, 1993). The idea that the grammar defines the language is abandoned in favour of a functional mapping, which assigns structures to input sentences, which even for an ideal grammar no longer need to be well-formed! Universal principles are used to restrict the space of admissible structures, where principles can be in contradiction to each other. There is, however, a strict ordering of principles available, defining the ability of a principle to override others. Thus, a complete disambiguation can always be achieved, at least for all the input sentences covered by the grammar.

In both cases the space of admissible structural descriptions is still limited by the context-free back-bone of the underlying rule set, which of course in theories like HPSG is very general and thus highly ambiguous during parsing. Constraint dependency grammar (CDG) (Maruyama, 1990) provides for an even more radical separation, by abandoning the notion of a rule altogether. The space of possible structural descriptions is defined completely independent from the grammar itself, which only consists of constraints on admissible structural descriptions. Furthermore, the approach also allows the handling of contradictory constraints (Menzel, 1998) by assigning weights to them, which describe their relative strength.

Since in a formal system which no longer maintains the consistency requirement, derivation amounts to determining one or several preferred interpretations from a huge set of competing ones, parsing becomes an optimisation problem. Instead of considering individual interpretations in terms of being consistent or not, alternative solutions have to be compared and a decision has to be made according to a given criterion (minimal extension, least severe constraint violation, maximal plausibility etc.). One of the most crucial questions, therefore, is to design a suitable computational framework which provides a good compromise between its expressive power on the one hand and the efficiency of the available solution procedures on the other. Initial experience has convinced us that CDGs with weighted constraints fulfill this requirement quite well and thus can be considered a good candidate even for practically feasible solutions.

4 Parsing with inconsistency - The CDG experience

Constraint dependency grammar is based on the idea of using constraint satisfaction techniques (Tsang, 1993) to disambiguate a packed dependency structure with a maximum degree of ambiguity. Initially the hypothesis space of such a disambiguation procedure contains all depencency relations between arbitrary pairs of input word forms. This situation is shown in Figure 1 for a simple German sentence and a set of dependency labels containing only three members. Even if only complete dependency trees are considered, 8000 different labelled trees can be extracted from this structure. Note that the size of

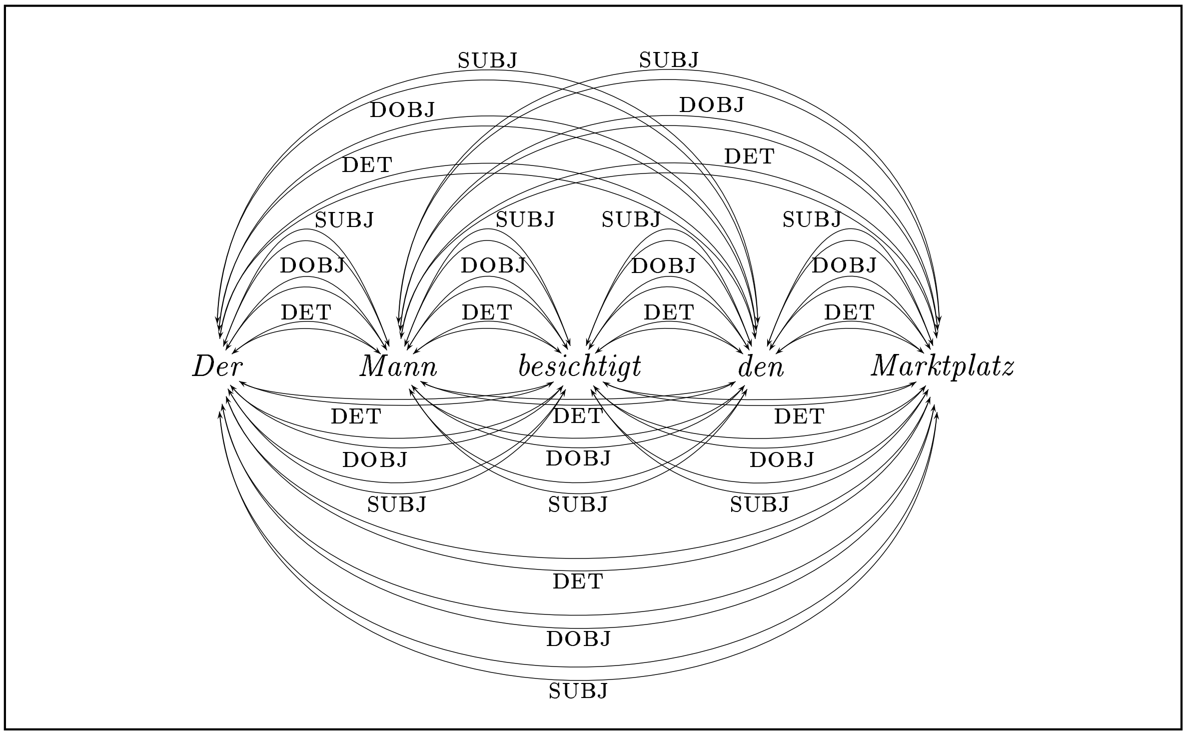

Figure 1: Initial state of a parsing problem with three labels (DET, SUBJ, DOBJ)

this hypothesis space only depends on the number of word forms and labels and therefore is almost completely independent of the grammar itself.

The grammar consists of a set of constraints, which describe the admissibility conditions for single dependency edges and pairs of them. Thus a great number of local interpretations can be removed from the space of hypotheses. In the example of Figure 1 three (highly simplified) unary constraints, namely

c_1: A determiner modifies a noun to the right.
c_2: A noun ist either subject or direct object of the finite verb.
c_3: A finite verb modifies nothing (i.e. is the root of the tree).

which consider only a single dependency edge at a time, already remove the vast majority of the initial possibilities. What they leave behind is the hypothesis space shown in Figure 2 with merely four trees remaining.

Two more binary constraints

c_4: A node cannot be modified twice by the same label.
c_5: The subject noun phrase has nominative case.

suffice to disambiguate the structure to the unique interpretation in Figure 3.

Although being universally quantified and therefore primarily suited to describe the possibility of a word form to serve as a dependant of another one, constraints can be used to model a limited number of existential conditions as well. They are needed to e.g. make the saturation of a valence requirement obligatory. For that purpose, additional auxiliary

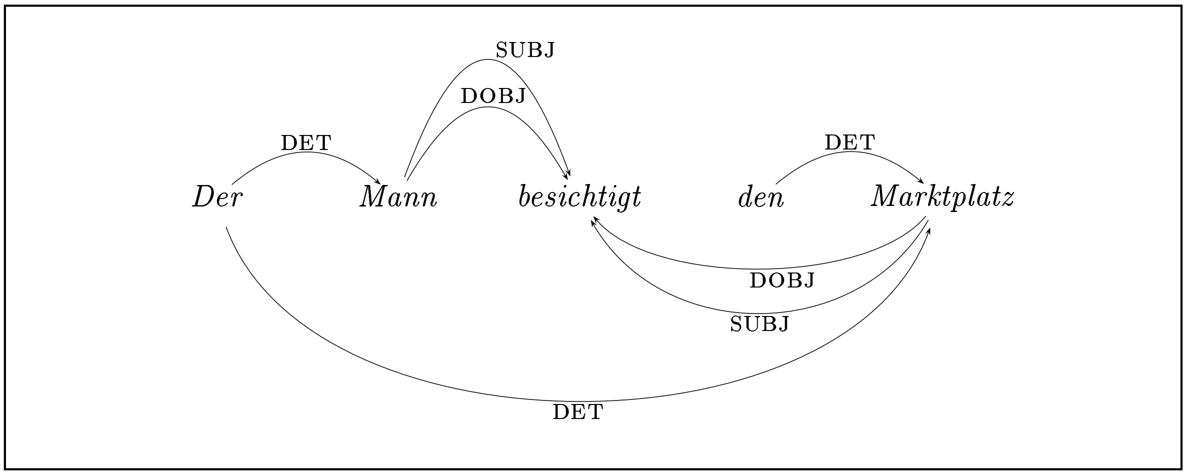

Figure 2: A partially disambiguated parsing problem.

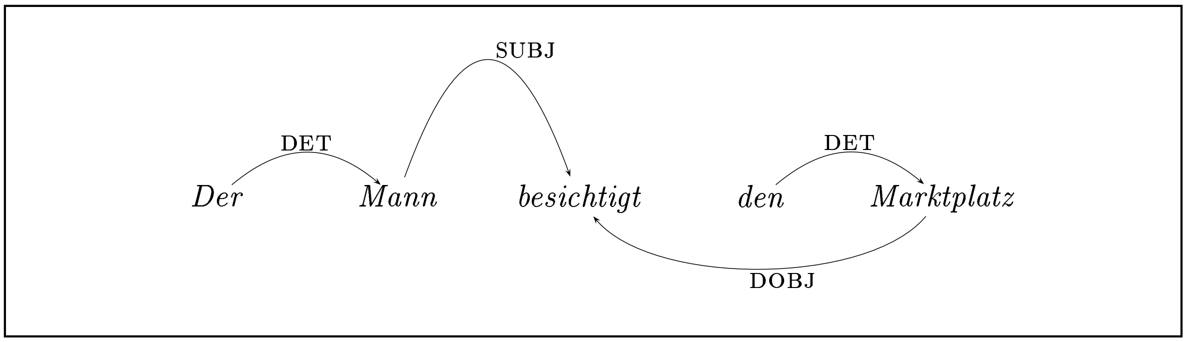

Figure 3: A completely disambiguated parsing problem.

levels of description are required, which mirror the original dependency relations, thus making e.g. a verb looking for a nominal modifier to which it can be attached. The idea of using additional description levels for special purposes can easily be extended to a multitude of representational layers, thus providing for a modular modelling of different structural aspects like syntax, semantics or informational structure. Figure 4 shows an example of such a multi-level representation, where only the syntactic relations are described as dependency tree above the sentence, whereas all the other levels, including thematic roles, a vorfeld slot and the auxiliary edges are drawn below.

To make this model work for real world parsing problems yet another extension becomes necessary. It concerns the handling of lexical ambiguity and, possibly, recognition uncertainty. This requires to generalise the disambiguation procedure from simple word sequences to word lattices (Harper and Helzermann, 1995). Unfortunately, this cannot be done without introducing an implicit global consistency criterion, requiring that all the readings which contribute to a solution must belong to a spanning path through the lattice.

Besides this, a fundamental problem remains. It is the notorious difficulty to devise a coherent set of constraints which is restrictive enough to achieve the desired disambiguation

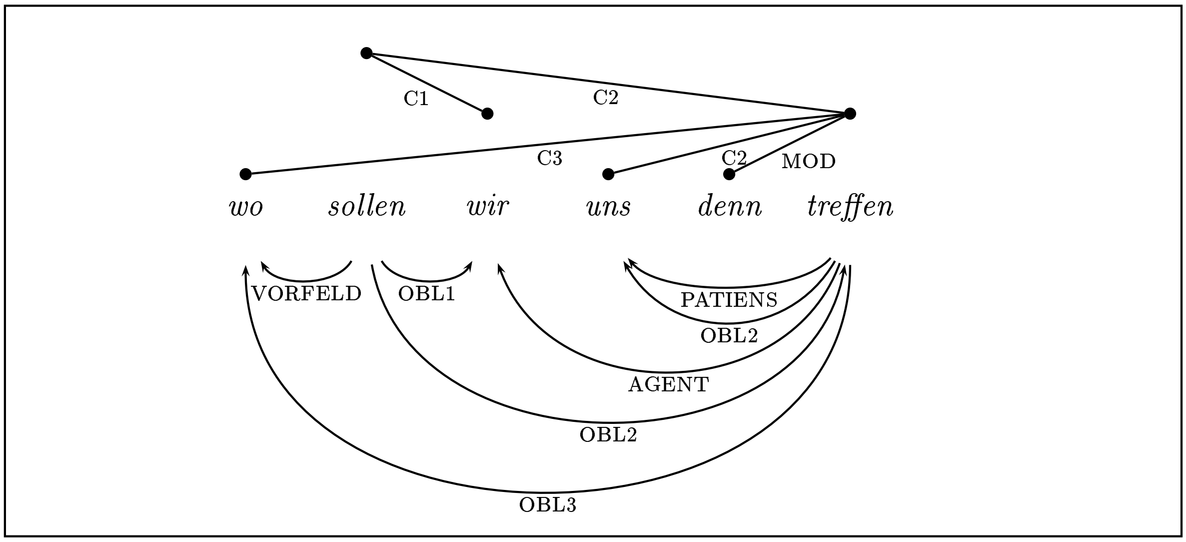

Figure 4: A multi-level representation combining aspects of syntactic, thematic and informational structure.

effect for arbitrary sentences, but at the same time permissive enough to not make the constraint satisfaction problem overconstrained (Harper et al., 1995). In the latter case the underlying logical theory becomes inconsistent and not solution can be obtained. Although a dynamic constraint invocation scheme could be devised, which prevents the last remaining structural interpretation from being discarded (Menzel, 1994), such an approach does not really address the underlying problem. Additionally, it would introduce a quite undesirable dependence on the order of constraint applications.

An alternative way to obtain solutions even for overconstrained (i.e. inconsistent) problems would be to use a scoring scheme for structural interpretations and to determine the best solution (or, if desired, the n-best ones) by means of an optimisation procedure. Scores can be based on the number or the kind of violated constraints. More general, however, seems to be an approach which assigns weights to constraints and combines them in an appropriate way. Conflicting evidence from a wide variety of information sources, including defaults and preferences, can then be arbitrated. Quite similar proposals have also been made for constraint-based unification grammars like HPSG (Uszkoreit, 1991). However, the complex nature of these constraints effectively seems to prevent efficient solution procedures from being devised for such a framework.

Thanks to the availability of a mechanism for evidence arbitration, contradictory requirements like those mentioned above can now be expressed in a convenient way. A number of solution methods to determine or approximate the desired optimum is available, among them techniques for guided search, heuristic pruning, and different kinds of transformation-based repair (Menzel and Schröder, 1998b; Foth, Menzel, and Schröder, 2000). Since they exhibit quite different properties with respect to soundness, completeness, convergence etc. they can be combined into a more general, concurrent architecture to take maximum advantage of their respective strengths.

The most striking observation, however, might be that thanks to the availability of a mechanism for evidence arbitration between model components solutions become possible which are quite attractive both from a cognitive *and* an application-oriented perspective:

1. Like in other soft-computing frameworks a certain degree of robustness against previously unseen data can be achieved, since inconsistencies between the model and the data can be resolved by the optimisation procedure and do not cause a breakdown of the parsing capabilities.

2. With their multi-level representation Constraint Dependency Grammars provide an integrated framework for combining evidence from very different knowledge sources while maintaining a strict modularity on the level of grammar and representation.

3. Due to the preference-driven combination of different informational contributions a qualitatively new level of robust behaviour can be reached, where knowledge deficits or even inconsistencies on one level can be compensated for by evidence from another one (Menzel, 1995). Most importantly this compensatory effect is bidirectional, so that syntactic deviations, caused e.g. by insufficient competence of the speaker, can be overcome relying on an appropriate semantic support, whereas syntactic evidence can override strange propositional content in case of e.g. novel, improbable or even false assertions.

4. The inherent robustness properties of the approach also extend to coping with temporal stress, where first experimental results indicate a graceful degradation under an increase of processing speed. Note that these two aspects of robustness can be traced back to the same combination of computational mechanisms.

5. Using the constraints which are violated by the optimum solution as diagnostic results about the utterance under consideration, anticipation-free systems for language learning can reach a qualitatively new performance level (Menzel and Schröder, 1998a; Menzel and Schröder, 1998c). Combined with other model-based techniques precise error explanations can be derived without restricting the freedom of choice of the student too much (Stockfleth, 2000).

6. The degree of robustness obtained can be exploited to analyse almost arbitrary input, including fragmentary sentences. This property can be used to pursue new ways of incremental parsing by considering successively extended sentence prefixes and trying to guide subsequent processing steps by the diagnostic findings about the currently available input (Foth et al., 2000).

Of course all these advantages do not come for free. Carrying out a comparison of alternative derivations, optimisation problems require *per se* much more computational effort than derivations in a comparable consistency-based model alone. Furthermore, the available results have been achieved at the expense of a limited expressive power of constraints: Efficiency reasons do not allow to consider more than two dependency edges at once. Note that optimality theory uses an even more severe restriction by constraining the domain of a principle to a single context-free rule. Further investigations in both areas will therefore have to substantiate the claim that large-coverage models of language can be built despite these limitations.

Compared to customary phrase structure approaches CDG profits considerably from the extremely regular structure of the hypothesis space of a dependency parsing problem, which is characterised by a unique branching factor and a unique depth. In particular CDG avoids the unlimited growth which (if not properly constrained) might result from the combination of context-free rules and the overparsing scheme proposed in (Tesar, 1995).

Despite their differences these independent moves towards more prowerful models of language seem to signal a kind of convergence between theory-driven and application-induced research goals. It gives rise to the optimism that we could expect a similar convergence on the side of methods and models as well. Already now we are witnessing how growing awareness of reality-inspired requirements for language modeling alongside with an appropriate choice of adequate abstractions brings linguistic theory in a position from which it might significantly contribute to more application-oriented research in Natural Language Processing. Abandoning the purely generative school of thought in favour of a more interpretative framework frees grammar from the strings of the consistency requirement and thus constitutes a major step towards the proper handling of real world data.

Doing so, however, requires to deal with a class of considerably more ambitious models, which hardly can be developed using traditional paper-and-pencil techniques alone. Hence, theoretical linguistics will certainly experience an increasing pressure to evaluate their results by means of ambitious implementations. This also might produce a strong incentive for closer cooperation links across our field.

5 Conclusions

There is good reason to assume that the striking separation between theoretically-driven research on the one hand and application-oriented work on the other has been caused by quite different assumptions about how the apparent inadequacy of one of the foundational cornerstones of linguistic modelling can be dealt with. Whereas many theoretical investigations continue to abstract away the problem of inconsistency by focussing on carefully selected subsets of language, application-oriented approaches, since being forced to handle all kind of naturally occuring data, resort to the collection of a constantly growing body of increasingly detailed special case descriptions. This not only is a labour-intensive approach, it additionally incurs severe processing problems as well.

Focussing on the example of the consistency requirement an attempt has been made to illustrate, how a fundamental shift in the fundamental abstractions adopted by a theory can pave a way towards a more realistic kind of linguistic models and thus offers a chance to overcome the lasting schism between theory and applications in Computational Linguistics.

References

Abney, Steven. 1997. Part-of-speech tagging and partial parsing. In G. Bloothooft and S. Young, editors, *Corpus-Based Methods in Language and Speech Processing*. Kluwer Academic Publisher, Dordrecht, chapter 4, pages 118–136.

Bod, Rens. 1995. *Enriching Linguistics with Statistics: Performance Models of Natural Language*. Ph.D. thesis, Department of Computational Linguistics, Universiteit van Amsterdam.

Brown, Peter F., John Cocke, Stephen A Della Pietra, Vincent J. Della Pietra, Frederick Jelinek, John D. Lafferty, Rober L. Mercer, and Paul S. Roossin. 1990. A statistical approach to machine translation. *Computational Linguistics*, 16(2):79–85.

Chelba, Ciprian and Frederick Jelinek. 1998. Exploiting syntactic structure for language modeling. In *Proceedings 17th International Conference on Computational Linguistics, 36th Annual Meeting of the ACL, Coling-ACL '98*, pages 225–231, Montreal, Canada.

Chomsky, Noam. 1957. *Syntactic Structures*. Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Clark, A. and R. Lutz. 1990. Guest-editorial-introduction. *AI & Society*, 4:3–16.

Douglas, Shona and Robert Dale. 1992. Towards robust PATR. In *Proceedings 14th International Conference on Computational Linguistics, Coling '92*, pages 468–474, Nancy, France.

Foth, Kilian, Wolfgang Menzel, Horia F. Pop, and Ingo Schröder. 2000. An experiment in incremental parsing using weighted constraints. In *Proceedings 18th International Conference on Computational Linguistics, Coling 2000*, Saarbrücken.

Foth, Kilian, Wolfgang Menzel, and Ingo Schröder. 2000. A transformation-based parsing technique with anytime properties. In *Proc. 4th Int. Workshop on Parsing Technologies*, pages 89–100, Trento, Italy.

Harper, Mary P. and Randall A. Helzermann. 1995. Extensions to constraint dependency parsing for spoken language processing. *Computer Speech and Language*, 9(3):187–234.

Harper, Mary P., Randall A. Helzermann, C. B. Zoltowski, B. L. Yeo, Y. Chan, T. Steward, and B. L. Pellom. 1995. Implementation issues in the development of the PARSEC parser. *Software - Practice and Experience*, 25(8):831–862.

Henderson, James and Peter Lane. 1998. A connectionist architecture for learning to parse. In *Proceedings 17th International Conference on Computational Linguistics, 36th Annual Meeting of the ACL, Coling-ACL '98*, pages 531–537, Montreal, Canada.

Heyer, G. 1988. Fünf Thesen zum kognitiven Anspruch und der industriellen Relevanz Künstlicher Intelligenz. In W. Hoeppner, editor, *Künstliche Intelligenz. GWAI-88, 12th German Workshop on Artificial Intelligence, Eringerfeld*. Springer, Berlin, Heidelberg, pages 320–326.

Heyer, G. 1989. Abstraktion und Allgemeinheit in sprachverarbeitenden Systemen im Spannungsfeld zwischen kognitiver Erklärungsadäquatheit und praxisorientierter Anwendung. In B. Becker, editor, *Zur Terminologie in der Kognitionsforschung (Workshop in der GMD)*. Gesellschaft für Mathematik und Datenverarbeitung mbH, Sankt Augustin, Germany, pages 194–209.

Kamp, Hans. 1981. A theory of truth and semantic representation. In J. Groenendijk, Th. Janssen, and M. Stokhof, editors, *Formal Method in the Study of Language*, Mathematisch Centrum Tracts. Amsterdam, pages 277–322.

Lavie, Alon. 1996. GLR*: A robust parser for spontaneously spoken language. In *ESSLLI-96 Workshop on Robust Parsing*, Prague.

Maruyama, H. 1990. Structural disambiguation with constraint propagation. In *Proceedings 28th Annual Meeting of the ACL*, pages 31–38.

Mellish, Chris. 1989. Some chart based techniques for parsing ill-formed input. In *Proceedings 27th Annual Meeting of the ACL*, pages 102–109, Vancouver.

Menzel, Wolfgang. 1994. Parsing of spoken language under time constraints. In T. Cohn, editor, *Proceedings 11th European Conference on Artificial Intelligence*, pages 560–564, Amsterdam.

Menzel, Wolfgang. 1995. Robust parsing of natural language. In *KI-95: Advances in Artificial Intelligence*, pages 19–34, Berlin. Springer-Verlag.

Menzel, Wolfgang. 1998. Constraint satisfaction for robust parsing of natural language. *Theoretical and Experimental Artificial Intelligence*, 10(1):77–89.

Menzel, Wolfgang and Ingo Schröder. 1998a. Constraint-based diagnosis for intelligent language tutoring systems. In *Proceedings IT & KNOWS, XV. IFIP World Computer Congress*, pages 484–497, Wien und Budapest.

Menzel, Wolfgang and Ingo Schröder. 1998b. Decision procedures for dependency parsing using graded constraints. In Sylvain Kahane and Alain Polguère, editors, *Proc. Coling-ACL Workshop on Processing of Dependency-based Grammars*, pages 78–87, Montreal, Canada.

Menzel, Wolfgang and Ingo Schröder. 1998c. Model-based diagnosis under structural uncertainty. In *Proceedings 13th European Conference on Artificial Intelligence*, pages 284–288, Brighton, UK.

Montague, Richard. 1974. *Formal Philosophy, Selected Papers of Richard Montague*. New Haven: Yale University Press.

Nagao, Makoto. 1988. Language engineering: The real bottleneck of natural language processing, panel with contributions by K. Jensen, D. Roesner, J. Tsuji and M. Tomita. In *Proceedings 12th International Conference on Computational Linguistics, Coling '88*, pages 448–453, Budapest, Hungary.

Prince, Alan and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Technical Report 2, Center for Cognitive Science, Rutgers University.

Shieber, Stuart M. 1992. *Constraint-Based Grammar Formalisms*. Cambridge, MA: Bradford.

Stöber, Karlheinz, Thomas Portele, Petra Wagner, and Wolfgang Hess. 1999. Synthesis by word concatenation. In *6th European Confence on Speech Communication and Technology*, pages 619–622, Budapest.

Stockfleth, Niels. 2000. Integration von CDG-Parsing und Fehlersimulation. Diplomarbeit, Universität Hamburg, Fachbereich Informatik.

Tesar, Bruce. 1995. *Computational Optimality Theory*. Ph.D. thesis, University of Colorado, Department of Computer Science, Boulder.

Tsang, E. 1993. *Foundations of Constraint Satisfaction*. Academic Press.

Uszkoreit, H. 1991. Strategies for adding control information to declarative grammars. In *Proceedings 29th Annual Meeting of the ACL*, pages 237–245.

Wahlster, Wolfgang, Thomas Bub, and Alex Waibel. 1997. Verbmobil: The combination of deep and shallow processing for spontaneous speech translation. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, pages 71–74, München.

Wermter, Stefan and Volker Weber. 1997. SCREEN: learning a flat syntactic and semantic spoken language analysis using artificial neural networks. *Journal of Artificial Intelligence Research*, 6(1):35–85.