Eliminative Parsing with Graded Constraints

Johannes Heinecke and **Jürgen Kunze**
(heinecke | kunze@compling.hu-berlin.de)
Lehrstuhl Computerlinguistik, Humboldt-Universität zu Berlin
Schützenstraße 21, 10099 Berlin, Germany

Wolfgang Menzel and **Ingo Schröder**
(menzel | ingo.schroeder@informatik.uni-hamburg.de)
Fachbereich Informatik, Universität Hamburg
Vogt-Kölln-Straße 30, 22527 Hamburg, Germany

Abstract

Natural language parsing is conceived to be a procedure of disambiguation, which successively reduces an initially totally ambiguous structural representation towards a single interpretation. Graded constraints are used as means to express well-formedness conditions of different strength and to decide which partial structures are locally least preferred and, hence, can be deleted. This approach facilitates a higher degree of robustness of the analysis, allows to introduce resource adaptivity into the parsing procedure, and exhibits a high potential for parallelization of the computation.

1 Introduction

Usually parsing is understood as a constructive process, which builds structural descriptions out of elementary building blocks. Alternatively, parsing can be considered a procedure of disambiguation which starts from a totally ambiguous structural representation containing all possible interpretations of a given input utterance. A combinatorial explosion is avoided by keeping ambiguity strictly local. Although particular readings can be extracted from this structure at every time point during disambiguation they are not maintained explicitly, and are not immediately available.

Ambiguity is reduced successively towards a single interpretation by deleting locally least preferred partial structural descriptions from the set of solutions. This reductionistic behavior coins the term *eliminative parsing*. The criteria which the deletion decisions are based on are formulated as compatibility constraints, thus parsing is considered a constraint satisfaction problem (CSP).

Eliminative parsing by itself shows some interesting advantages:

Fail soft behavior: A rudimentary robustness can be achieved by using procedures that leave the last local possibility untouched. More elaborated procedures taken from the field of partial constraint satisfaction (PCSP) allow for even greater robustness (cf. Section 3).

Resource adaptivity: Because the sets of structural possibilities are maintained explicitly, the amount of disambiguation already done and the amount of the remaining effort are immediately available. Therefore, eliminative approaches lend themselves to the active control of the procedures in order to fulfill external resource limitations.

Parallelization: Eliminative parsing holds a high potential for parallelization because ambiguity is represented locally and all decisions are based on local information.

Unfortunately even for sublanguages of fairly modest size in many cases no complete disambiguation can be achieved (Harper et al., 1995). This is mainly due to the crisp nature of classical constraints that do not allow to express the different strength of grammatical conditions: A constraint can only allow or forbid a given structural configuration and all constraints are of equal importance.

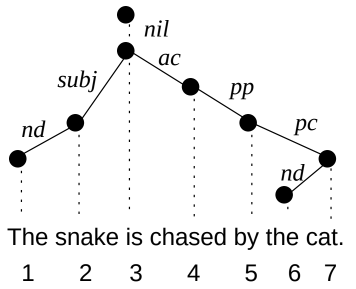
To overcome this disadvantage gradings can be added to the constraints. Grades indicate how serious one considers a specific constraint violation and allow to express a range of different types of conditions including preferences, defaults, and strict restrictions. Parsing, then, is modelled as a partial constraint satisfaction problem with scores (Tsang, 1993) which can almost always be disambiguated towards a single solution if only the grammar provides enough evidence, which means that the CSP is overconstrained in the classical sense because at least preferential constraints are violated by the solution.

We will give a more detailed introduction to constraint parsing in Section 2 and to the extension to graded constraints in Section 3. Section 4 presents algorithms for the solution of the previously defined parsing problem and the linguistic modeling for constraint parsing is finally described in Section 5.

2 Parsing as Constraint Satisfaction

While eliminative approaches are quite customary for part-of-speech disambiguation (Padró, 1996) and underspecified structural representations (Karlsson, 1990), it has hardly been used as a basis for full structural interpretation. Maruyama (1990) describes full parsing by means of constraint satisfaction for the first time.

(a)



(b)

$$\begin{array}{lllllll} v_1 & = & \langle nd, 2\rangle & v_2 & = & \langle subj, 3\rangle \\ v_3 & = & \langle nil, 0\rangle & v_4 & = & \langle ac, 3\rangle \\ v_5 & = & \langle pp, 4\rangle & v_6 & = & \langle nd, 7\rangle \\ v_7 & = & \langle pc, 5\rangle & & & \end{array}$$

Figure 1: (a) Syntactic dependency tree for an example utterance: For each word form an unambiguous subordination and a label, which characterizes of subordination, are to be found. (b) Labellings for a set of constraint variables: Each variable corresponds to a word form and takes a pairing consisting of a label and a word form as a value.

Dependency relations are used to represent the structural decomposition of natural language utterances (cf. Figure 1a). By not requiring the introduction of non-terminals, dependency structures allow to determine the initial space of subordination possibilities in a straight forward manner. All word forms of the sentence can be regarded as constraint variables and the possible values of these variables describe the possible subordination relations of the word forms. Initially, all pairings of a possible dominating word form and a label describing the kind of relation between dominating and dominated word form are considered as potential value assignments for a variable. Disambiguation, then, reduces the set of values until finally a unique value has been obtained for each variable. Figure 1b shows such a final assignment which corresponds to the dependency tree in Figure 1a.[1]

Constraints like

$\{\mathsf{X}\}$: Subj : Agreement : X.label=subj $\rightarrow$
X↓cat=NOUN $\wedge$ X↑cat=VERB $\wedge$ X↓num=X↑num

judge the well-formedness of combinations of subordination edges by considering the lexical properties of the subordinated (X↓num) and the dominating (X↑num) word forms, the linear precedence (X↑pos) and the labels (X.label). Therefore, the conditions are stated on structural representations rather than on input strings directly. For instance, the above constraint can be paraphrased as follows: Every subordination as a subject requires a noun to be subordinated and a verb as the dominating word form which have to agree with respect to number.

An interesting property of the eliminative approach is that it allows to treat unexpected input without the necessity to provide an appropriate rule beforehand: If constraints do not exclude a solution explicitly it will be accepted. Therefore, defaults for unseen phenomena can be incorporated without additional effort. Again there is an obvious contrast to constructive methods which are not able to establish a structural description if a corresponding rule is not available.

For computational reasons only unary and binary constraints are considered, i. e. constraints interrelate at most two dependency relations. This, certainly, is a rather strong restriction. It puts severe limitations on the kind of conditions one wishes to model (cf. Section 5 for examples). As an intermediate solution, templates for the approximation of ternary constraints have been developed.

Harper et al. (1994) extended constraint parsing to the analysis of word lattices instead of linear sequences of words. This provides not only a reasonable interface to state-of-the-art speech recognizers but is also required to properly treat lexical ambiguities.

3 Graded Constraints

Constraint parsing introduced so far faces at least two problems which are closely related to each other and cannot easily be reconciled. On the one hand, there is the difficulty to reduce the ambiguity to a single interpretation. In terms of CSP, the constraint parsing problem is said to have too small a tightness, i. e. there usually is more than one solution. Certainly, the remaining ambiguity can be further reduced by adding additional constraints. This, on the other hand, will most probably exclude other constructions from being handled properly, because highly restrictive constraint sets can easily render a problem unsolvable and therefore introduce brittleness into the parsing procedure. Whenever being faced with such an overconstrained problem, the procedure has to retract certain constraints in order to avoid the deletion of indispensable subordination possibilities.

Obviously, there is a trade-off between the coverage of the grammar and the ability to perform the disambiguation efficiently. To overcome this problem one wishes to specify exactly *which* constraints can be relaxed in case a solution can not be established otherwise. Therefore, different types of con-

[1]For illustration purposes, the position indices serve as a means for the identification of the word forms. A value $\langle nil, 0\rangle$ is used to indicate the root of the dependency tree.

straints are needed in order to express the different strength of strict conditions, default values, and preferences.

For this purpose every constraint c is annotated with a weight $w(c)$ taken from the interval $[0, 1]$ that denotes how seriously a violation of this constraint effects the acceptability of an utterance (cf. Figure 2).

{X} : SubjInit : Subj : 0.0 :
X.label=subj $\rightarrow$ X$\downarrow$cat=NOUN $\wedge$ X$\uparrow$cat=VERB
{X} : SubjNumber : Subj : 0.1 :
X.label=subj $\rightarrow$ X$\downarrow$num=X$\uparrow$num
{X} : SubjOrder : Subj : 0.9 :
X.label=subj $\rightarrow$ X$\downarrow$pos<X$\uparrow$pos
{X, Y} : SubjUnique : Subj : 0.0 :
X.label=subj $\wedge$ X$\uparrow$id=Y$\uparrow$id $\rightarrow$ Y.label$\neq$subj

Figure 2: Very restrictive constraint grammar fragment for subject treatment in German: Graded constraints are additionally annotated with a score.

The solution of such a *partial constraint satisfaction problem with scores* is the dependency structure of the utterance that violates the fewest and the weakest constraints. For this purpose the notation of constraint weights is extended to scores for dependency structures. The scores of all constraints c violated by the structure under consideration s are multiplied and a maximum selection is carried out to find the solution s' of the PCSP.

$$s' = \arg\max_{s} \prod_{c} w(c)^{n(c,s)}$$

Since a particular constraint can be violated more than once by a given structure, the constraint grade $w(c)$ is raised to the power of $n(c, s)$ which denotes the number of violations of the constraint c by the structure s.

Different types of conditions can easily be expressed with graded constraints:

- Hard constraints with a score of zero (e. g. constraint SubjUnique) exclude totally unacceptable structures from consideration. This kind of constraints can also be used to initialize the space of potential solutions (e. g. SubjInit).
- Typical well-formedness conditions like agreement or word order are specified by means of weaker constraints with score larger than, but near to zero, e. g. constraint SubjNumber.
- Weak constraints with score near to one can be used for conditions that are merely preferences rather than error conditions or that encode uncertain information. Some of the phenomena one wishes to express as preferences concern word order (in German, cf. subject topicalization of constraint SubjOrder), defeasible selectional restrictions, attachment preferences, attachment defaults (esp. for partial parsing), mapping preferences, and frequency phenomena. Uncertain information taken from prosodic clues, graded knowledge (e. g. measure of physical proximity) or uncertain domain knowledge is a typical example for the second type.

Since a solution to a CSP with graded constraints does not have to satisfy every single condition, overconstrained problems are no longer unsolvable. Moreover, by deliberately specifying a variety of preferences nearly all parsing problems indeed become overconstrained now, i. e. no solution fulfills all constraints. Therefore, disambiguation to a single interpretation (or at least a very small solution set) comes out of the procedure without additional effort. This is also true for utterances that are — strictly speaking — grammatically ambiguous. As long as there is any kind of preference either from linguistic or extra-linguistic sources no enumeration of possible solutions will be generated.

Note that this is exactly what is required in most applications because subsequent processing stages usually need only one interpretation rather than many. If under special circumstances more than one interpretation of an utterance is requested this kind of information can be provided by defining a threshold on the range of admissible scores.

The capability to rate constraint violations enables the grammar writer to incorporate knowledge of different kind (e. g. prosodic, syntactic, semantic, domain-specific clues) without depending on the general validity of every single condition. Instead, occasional violations can be accepted as long as a particular source of knowledge supports the analysis process in the long term.

Different representational levels can be established in order to model the relative autonomy of syntax, semantics, and even other contributions. These multiple levels must be related to each other by means of mapping constraints so that evidence from one level helps to find a matching interpretation on another one. Since these constraints are defeasible as well, an inconsistency among different levels must not necessarily lead to an overall break down.

In order to accommodate a number of representational levels the constraint parsing approach has to be modified again so that a separate constraint variable is established for each level and each word form. A solution, then, does not consist of a single dependency tree but a whole set of trees.

While constraint grades make it possible to weigh up different violations of grammatical conditions the representation of different levels additionally allows for the arbitration among conflicting evidence originating from very different sources, e. g. among agreement conditions and selectional role filler restrictions or word order regularities and prosodic hints.

While constraints encoding specific domain knowledge have to be exchanged when one switches to another application context other constraint clusters like syntax can be kept. Consequently, the multi-level approach which makes the origin of different disambiguating information explicit holds great potential for reusability of knowledge.

4 Solution methods

In general, CSPs are NP-complete problems. A lot of methods have been developed, though, to allow for a reasonable complexity in most practical cases. Some heuristic methods, for instance, try to arrive at a solution more efficiently at the expense of giving up the property of correctness, i. e. they find the globally best solution in most cases while they are not guaranteed to do so in all cases. This allows to influence the temporal characteristics of the parsing procedure, a possibility which seems especially important in interactive applications: If the system has to deliver a reasonable solution within a specific time interval a dynamic scheduling of computational resources depending on the remaining ambiguity and available time is necessary (Menzel, 1994, *anytime algorithm*). While different kinds of search are more suitable with regard to the correctness property, local pruning strategies lend themselves to resource adaptive procedures. Menzel and Schröder (1998b) give details about the decision procedures for constraint parsing.

5 Grammar modeling

For experimental purposes a constraint grammar has been set up, which consists of two descriptive levels, one for syntactic (including morphology and agreement) and one for semantic relations. Whereas the syntactical description clearly follows a dependency approach, the second main level of our analysis, semantics, is limited to sortal restrictions and predicate-argument relations for verbs, predicative adjectives, and predicative nouns.

In order to illustrate the interaction of syntactical and semantical constraints, the following (syntactically correct) sentence is analyzed. Here the use of a semantic level excludes or depreciates a reading which violates lexical restrictions: *Da habe ich einen Termin beim Zahnarzt* ("At this time, I have an appointment at the dentist's.") The preposition *beim* ("at the") is a locational preposition, the noun *Zahnarzt* ("dentist"), however, is of the sort "human". Thus, the constraint which determines sortal compatibility for prepositions and nouns is violated:

{X} : PrepSortal : Prepositions : 0.3 :
X↑cat=PREP X↓cat=NOUN →
compatible(ont, X↑sort, X↓sort)
'Prepositions should agree sortally with their noun.'

Other constraints control attachment preferences. For instance, the sentence *am Montag machen wir einen Termin aus* has two different readings ("we will make an appointment, which will take place on Monday" vs. "on Monday we will meet to make an appointment for another day"), i. e. the attachment of the prepositional phrase *am Montag* can not be determined without a context. If the first reading is preferred (the prepositional phrase is attached to *ausmachen*), this can be achieved by a graded constraint. It can be overruled, if other features rule out this possibility.

A third possible use for weak constraints are attachment defaults, if e. g. a head word needs a certain type of word as a dependent constituent. Whenever the sentence being parsed does not provide the required constituent, the weak constraint is violated and another constituent takes over the function of the "missing" one (e. g. nominal use of adjectives).

Prosodic information could also be dealt with. Compare *Wir müssen noch einen Termin ausmachen* ("We *still* have to make an appointment" vs. "We have to make a *further* appointment"). A stress on *Termin* would result in a preference of the first reading whereas a stressed *noch* makes the second translation more adequate. Note that it should always be possible to outdo weak evidence like prosodic hints by rules of word order or even information taken from the discourse, e. g. if there is no previous appointment in the discourse.

In addition to the two main description levels a number of auxiliary ones is employed to circumvent some shortcomings of the constraint-based approach. Recall that the CSP has been defined as to uniquely assign a dominating node (together with an appropriate label) to each input form (cf. Figure 1). Unfortunately, this definition restricts the approach to a class of comparatively weak well-formedness conditions, namely *subordination possibilities* describing the degree to which a node can fill the valency of another one. For instance, the potential of a noun to serve as the grammatical subject of the finite verb (cf. Figure 2) belongs to this class of conditions. If, on the other hand, the somewhat stronger notion of a *subordination necessity* (i. e. the requirement to fill a certain valency) is considered, an additional mechanism has to be introduced. From a logical viewpoint, constraints in a CSP are universally quantified and do not provide a natural way to accomodate conditions of existence. However, in the case of subordination necessities the effect of an existential quantifier can easily be simulated by the unique value assignment principle of the constraint satisfaction mechanism itself. For that purpose an additional representational level for the *inverse* dependency relation is introduced for each valency to be saturated (Helzerman and Harper, 1992, cf. needs-roles). Dedicated constraints ensure that the inverse relation can only be established if a suitable filler has properly been identified in the input sentence.

Another reason to introduce additional auxiliary levels might be the desire to use a feature inheritance mechanism within the structural description. Basically, constraints allow only a passive feature checking but do not support the active assignment of feature values to particular nodes in the dependency tree. Although this restriction must be considered a fundamental prerequisite for the strictly local treatment of huge amounts of ambiguity, it certainly makes an adequate modelling of feature percolation phenomena rather difficult. Again, the use of auxiliary levels provides a solution by allowing to transport the required information along the edges of the dependency tree by means of appropriately defined labels. For efficiency reasons (the complexity is exponential in the number of features to percolate over the same edge) the application of this technique should be restricted to a few carefully selected phenomena.

The approach presented in this paper has been tested successfully on some 500 sentences of the Verbmobil domain (Wahlster, 1993). Currently, there are about 210 semantic constraints, including constraints on auxiliary levels. The syntax is defined by 240 constraints. Experiments with slightly distorted sentences resulted in correct structural trees in most cases.

6 Conclusion

An approach to the parsing of dependency structures has been presented, which is based on the elimination of partial structural interpretations by means of constraint satisfaction techniques. Due to the graded nature of constraints (possibly conflicting) evidence from a wide variety of informational sources can be integrated into a uniform computational mechanism. A high degree of robustness is introduced, which allows the parsing procedure to compensate local constraint violations and to resort to at least partial interpretations if necessary.

The approach already has been successfully applied to a diagnosis task in foreign language learning environments (Menzel and Schröder, 1998a). Further investigations are prepared to study the temporal characteristics of the procedure in more detail. A system is aimed at, which eventually will be able to adapt its behavior to external pressure of time.

Acknowledgements

This research has been partly funded by the German Research Foundation "Deutsche Forschungsgemeinschaft" under grant no. Me 1472/1-1 & Ku 811/3-1.

References

Mary P. Harper, L. H. Jamieson, C. D. Mitchell, G. Ying, S. Potisuk, P. N. Srinivasan, R. Chen, C. B. Zoltowski, L. L. McPheters, B. Pellom, and R. A. Helzerman. 1994. Integrating language models with speech recognition. In *Proceedings of the AAAI-94 Workshop on the Integration of Natural Language and Speech Processing*, pages 139–146.

Mary P. Harper, 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.

Randall A. Helzerman and Mary P. Harper. 1992. Log time parsing on the MasPar MP–1. In *Proceedings of the 6th International Conference on Parallel Processing*, pages 209–217.

Fred Karlsson. 1990. Constraint grammar as a framework for parsing running text. In *Proceedings of the 13th International Conference on Computational Linguistics*, pages 168–173, Helsinki.

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

Wolfgang Menzel and Ingo Schröder. 1998a. Constraint-based diagnosis for intelligent language tutoring systems. In *Proceedings of the IT&KNOWS Conference at the IFIP '98 Congress*, Wien/Budapest.

Wolfgang Menzel and Ingo Schröder. 1998b. Decision procedures for dependency parsing using graded constraints. In *Proc. of the Joint Conference COLING/ACL Workshop: Processing of Dependency-based Grammars*, Montreal, CA.

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

Lluís Padró. 1996. A constraint satisfaction alternative to POS tagging. In *Proc. NLP+IA*, pages 197–203, Moncton, Canada.

E. Tsang. 1993. *Foundations of Constraint Satisfaction.* Academic Press, Harcort Brace and Company, London.

Wolfgang Wahlster. 1993. Verbmobil: Translation of face–to–face dialogs. In *Proceedings of the Machine Translation Summit IV*, pages 127–135, Kobe.