Verbmobil
Verbundvorhaben

Morphological Earley-based Chart Parsing in Connected Word Recognition

Martina Pampel

Universität Bielefeld

Report 152
August 1996

August 1996

Martina Pampel
Universität Bielefeld (UBI)
Fakultät für Linguistik und Literaturwissenschaft
Universitätsstr. 25
Postfach 10 01 31
33501 Bielefeld

Tel.: (0521) 106 - 3510
Fax: (0521) 106 - 6008
e-mail: {martina}@Spectrum.Uni-Bielefeld.DE

Gehört zum Antragsabschnitt: 15.6 Interaktive Phonologische Interpretation

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 101 B 2 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei dem Autor.

1 Goals and Motivation

One of the most disputed aspects of automatic speech recognition (ASR) has traditionally been the role of linguistic knowledge versus statistical corpus-based training; recently however, the viewpoints have been increasingly converging (4). In this paper the claim is made that computational linguistic techniques can contribute towards the solution of several difficult problems which are currently under discussion in connection with word recognition in spontaneous continuous speech: (1) The recognition of 'unknown' but well-formed words; (2) Reduction of the noisiness of statistical language models in highly inflecting languages; (3) Improvement of performance by the use of sub-units based on morphological structure; (4) Support of new incremental, interactive speech recognition algorithms with 'anytime' properties; (4) Mapping of temporal signal sample lattices on to abstract symbolic chart structures.

An experimental workbench for linguistically based word recognition in German was developed, based on the premise that these problems are related, which involves three unusual features for ASR: (1) a detailed morphological word syntax for spoken German, (2) word syntax categories based on underspecified feature structures (underspecified phonemes); (3) a parametrised Earley-type lattice-to-chart parser for lattices of underspecified phonemes. In the parser, the PREDICT component is parametrised for different interactive prediction types, the SCAN component is enhanced for the lattice-to-chart mapping problem, and the COMPLETE component is modified to account for competing hypotheses at unconnected chart nodes with temporal offsets relative to each other. The solution has been implemented as a component of a linguistically based word recogniser for continuous spontaneous speech within the context of a spoken language translation system.

Bottom up chart parsing technology has been used in ASR at least since 1983 (2); (8), and suggestions for non-incremental lattice parsing at sentence level have been made by (1). The selection of modern active chart parser (ACP) technology (11) for word parsing is based on the criteria of *efficiency*, *monotonicity*, *incrementality*, *interactivity*, and *anytime properties*:

- *Efficiency.* Intermediate alternative parse results are stored as items in an optimised graph in the chart data structure.
- *Monotonicity.* The chart data structure as a monotonically constructed graph of well-formed substrings, together with breadth-first search, enables use of constraint oriented inference techniques for multiple knowledge sources.

- *Flexible inference.* Access to items of the chart is not rigidly constrained by the data structure. Various search strategies can be used to handle the order of access and it is imaginable that search strategies can be varied within the analysis.

- *Incrementality.* Left-right control regimes over chart nodes model temporal precedence flow expressed in temporal phoneme lattices and permits handling of partial (underspecified) analyses.

- *Interactivity.* The PREDICT component of an ACP can be parametrised to include top-down search constraints from other knowledge sources.

- *Anytime properties.* The monotonicity and incrementality properties mean that an ACP allows retrieval of all available and even partial results from the chart and is therefore intrinsically suitable for use as an anytime algorithm (5); (7) and for incorporation in incremental, interactive ASR architectures (10).

2 Morphology in automatic speech recognition

Connected word recognition in current speech technology is often done on the basis of stochastic modelling with Hidden Markov Modells (HMMs), a form of probabilistic finite automaton, or with neural networks (NNs). The main problems with these approaches are that it is not clear how to use HMMs and NNs in incremental interactive architectures (12), moreover it is not able to handle unknown words properly since recognition performance depends on the use of corpus training data for modelling phonological segments (phonemes) and a corpus-based language model for modelling constraints on word sequences. Though well-formedness constraints on linguistic units like phonemes, syllables, morphemes and words can be implicitly modelled in a HMM, no explicit generalisations about those units as in a symbolic linguistic approach exist (but see (4)), and a fixed control regime is used.

The use of morphology in modern word recognition architectures is well- motivated. Firstly, the linguistic word sub-units used in ASR can be linguistically characterised as constituents of the lower part of a prosodic hierarchy: phonemes, semi-syllables, syllables, phonological words, and the segmentation of words into such units does not correspond exactly to a segmentation into morphological units; morphological units and morphological constraints are, however, required for the integration of syntactic top-down information (for example for specifying inflectional properties of words).

Secondly, arguments against an integration of linguistic constraints in word recognition applications were often made in connection with English, a language with very few inflections. However, ASR results for such a language can obviously not be transferred to highly inflecting languages, where - given the same number of word stems - the vocabulary of fully inflected forms is larger by a factor > 3 (German) or > 1000 (Finnish; (6). The claim is made here that the superiority of whole word statistics over symbolic knowledge does not necessarily hold for these highly inflecting or agglutinative languages, and technology based on unanalysed whole word forms, as used in HMMs needs to be augmented by technology based on morphological analysis.

Thirdly, morphological analysis and morphological generalisations provide techniques for defining patterns for new words, as well as providing compression techniques for reducing lexicon size.

3 Morphology and the recognition of new words

The distinction between actual and potential linguistic units can be made at all linguistic levels. Potential units are not listed in the lexicon, the moment they are invented but are well-formed according to the rules of the grammar (phonotactic, morphotactic, syntax).

The German prefix “un-” is very productive for negation purposes; the resulting words may be lexicalised or freely constructed. Derivations like “unartig”, “unhöflich”, “unsauber”, may be lexicalised, but examples such as “uncool”, “ungefaxt” need not be stored as lexical entries but can be derived with a morphotactic rule (13).

New words based on loan words are frequent in languages such as German. Words like German “gefaxt” /g@fakst/, “angemailt” /ang@me:lt/, “geprunt” /g@pru:nt/ are examples of a productive process in which a German verb is built by concatenation of the stem of an English verb with the German inflection schema for inseparable, regular verbs. Furthermore it is evidently impossible to store numbers in large or non-finite domains (e.g. the results of arithmetic operations) in a lexicon; morphological “number syntax” rules are needed.

Finally, *ad hoc* compound and derived words are frequent in continuous speech, and need to be handled by morphotactic rules. New technical terms have the same status.

4 Morphological parsing in word recognition

Feature-based word grammars are a possible solution to the problem of hypotheses explosion. Feature structures permit the processing of partial derivations of underspecified words because the basic units in a feature-based grammar are categories. The underspecification of categories is possible whereas in a pure context-free grammar the applicability of a rule to a set of labels is defined through the identity operation and no degrees of similarity of labels can be introduced. In a feature grammar one can have partial or underspecified categories within rules and each category in a rule is deemed to apply to any constituent whose label is a superset of that category (14). The morphological component presented in (15) uses such a feature-based morphotactic grammar, a morph and a lexeme lexicon, but the input format in the interpreting mode is specified in form of isolated written words and is therefore not applicable, as it stands, to spoken language parsing. Also, the interpreter does not handle temporal lattice parsing.

The division between word recognition and syntactic parsing in most of the current speech understanding systems is generally seen as a division between stochastically oriented acoustics and symbolically oriented linguistics, with the consequence that modern computational linguistic data structures are not used below the word level, and the interface between acoustics and linguistics is defined as a string-based word hypothesis graph, often in orthographic rather than phonemic representation which might be expected (9).

From a linguistic point of view two improvements are possible here. Firstly, use of a phonologically oriented feature-based interface between word recognition and sentence parsing makes it possible to include analyses with finer granularity. Secondly, processing at this interface can supported by a morphological component; together with a syntactic component it can contribute towards a drastic reduction of the size of the word hypothesis lattice. Thirdly, the endings of words cannot be detected reliably by the acoustic component, and recognition of inflected items is thus noisy; from a morphological point of view, only underspecified information about stems is reliably available. While a stochastic model has to map the noisy information to all stored word models and has to send a disjunction of the most suitable fully specified word hypotheses to a syntactic component, a word recognition component with morphotactic knowledge promises noise reduction at this point.

The ACP-based morphological component for spoken language presented in this paper has the following features:

- interfaces to a phonological and a syntactic component.

- Input: a lattice of underspecified feature-based phonological hypothesis labelled with their signal boundary points in msec. The phonological features are phonation, manner, place, length, height, roundness and openness.

- The morphologically structured output to the syntactic component is currently mapped to a word lattice with complete word forms for conventional evaluation purposes.
- For test purposes the morphotactic feature grammar is implemented as a Definite Clause Grammar with chart parsing. The implementation is based on (3).

5 Active chart parser modifications

The modifications of the three operations PREDICT, SCAN and COMPLETE contribute to relax the constraints embodied in the operations. The following definitions are required for discussion: A formal context-free grammar G is a quadruple $G = <\mathrm{V_T}, \mathrm{V_N}, S, R>$, where

- V_{T} and V_{N} are nonempty finite disjoint sets,
- S is a distinguished member or a set of V_{N},
- R is a finite set of ordered pairs in $V_{\mathrm{N}} \times$ Z.

Let Z $= V_{\mathrm{T}} \cup V_{\mathrm{N}}$. In rule notation, the element of V_{N} is the left hand side (LHS) and the elements in Z are the right hand side of the rule: LHS $\rightarrow$ RHS. $\mathrm{R_s}$ is a subset of R for which holds that LHS belongs to the category S. Let A be an element of V_{N} and a of V_{T}. An *item* represents the state of the application of a grammar rule in R and is a kind of record with at least five arguments:

$$[BEGIN, END, CAT, ToPARSE, PARSED]_{\mathrm{Item}} \quad ,$$

where

- The arguments BEGIN and END are tuples of the form $< BP, CN >$ of a boundary point BP and a related chart node CN;
- BEGIN denotes the beginning point and
- END the ending point of the hypothesis. This is an extension to Earley's definition of items relevant for speech recognition applications. The argument
- CAT denotes the category of the item.

- The argument PARSED denotes input which is already parsed. The argument ToPARSE denotes expectations.

The chart items are defined as follows:

- An inactive terminal item ($I_{\mathrm{t\_inact}}$) has the form [BEGIN,END,a,[],PARSED]. It contains a bottom-up hypothesis (bu-hypo).

- An inactive nonterminal item ($I_{\mathrm{nont\_inact}}$) has the form BEGIN,END,A,[],PARSED], and an inactive s-item ($I_{\mathrm{s\_inact}}$) has the form [BEGIN,END,S,[],PARSED].

- An active nonterminal item ($I_{\mathrm{nont\_act}}$) has the form BEGIN,END,ToPARSE,PARSED].

- A cyclic item (I_{c}) has the form [X,X,ToPARSE,[]] It has identical beginning and ending points. Cyclic items are generated in the predict operation and correspond to td-hypos about future input. In general those items do not have temporal extensions. Items representing top-down hypotheses (td-hypo) might include stochastic statements about the average temporal extension of the hypothesis and confidence values for the hypothesis.

5.1 The predictor

The predict operation of Earley's algorithm is parametrised into different predict modes in order to test different search spaces for bu-hypos and examine resulting processing speeds. A first distinction has to be made between initial and normal predict operations:

- INITIAL PREDICT: For all grammar rules of the set $\mathrm{R_s}$ LHS $\rightarrow$ RHS and for all grammar rules which have a left hand side category which is unifiable with one of the left corner categories of the rules in $\mathrm{R_s}$ items of type $\mathrm{I_c}$ will be generated.

- NORMAL PREDICT: The task of this operation is to calculate all expectations of an active item added to the chart and to store these expectations as items $\mathrm{I_c}$ in the chart.

The following parametrisation of the predict mode concerns the application of these two predict versions to define variant operations.

- STANDARD PREDICT: This predict mode is the one specified in Earley's algorithm and involves an INITIAL PREDICT and a NORMAL PREDICT. In the beginning a single INITIAL PREDICT is performed for the start category of an item and a NORMAL PREDICT is executed every time an active item has been added to the chart. Under a STANDARD PREDICT the parser terminates if one unit of the category S is recognized or if no more input is available. This mode is only suitable for written language applications with a predefined category S for the whole input.

- SOMEWHERE PREDICT: An INITIAL PREDICT is done every time an item I_{s_inact} is found. This means that for every node which is an ending node of an item I_{s_inact} stored in the chart the INITIAL PREDICT operation is executed. This mode allows for continuous speech recognition tasks where the input can be devided in several substrings of category S. Under a SOMEWHERE PREDICT the parser terminates if no more input is available. If at any processing stage no more units of category S can be recognized up to this point the parser only adds inactive terminal items to the chart which cannot be completed anymore even if they fulfil the premises of the completor. This is because no more INITIAL PREDICT operations will be made. This mode is only suitable for well-formed input according to the rules of grammar.

- EVERYWHERE PREDICT: An INITIAL PREDICT is done for the starting node and for every node which is an ending node of a scanned input hypothesis. This mode takes into consideration the characteristics of continuous speech like non well-formed inputs, breaking offs, fragments, uncertain beginnings and endings of linguistic units. Under an EVERYWHERE PREDICT the parser terminates if no more input is available and consequently gives the possibility to recognize all units of category S. The advantage of this mode is that it can handle uncertain beginnings and endings of words, fragments, etc. The disadvantage of this mode is that it builds up a large search space.

- ANYWHERE PREDICT: This predict mode corresponds to the behaviour of a pure bottom-up chart parser. The INITIAL PREDICT is modified with respect to the kind of categories S. All categories of the grammar are specified as start categories in this mode and therefore all grammar rules are used for the generation of cyclic items. The advantage of this mode is that it catches the complete analysis. The disadvantage of this mode is the same as in EVERYWHERE PREDICT.

The available top-down strategy in standard ACP algorithms can be used for the interaction with other components. Therefore we have to distinguish between different categories of td-hypo:

- external td-hypos, namely hypothesis which come from other components,
- internal td-hypos which correspond to predictions based on components internal grammars as discussed above.

The td-hypos restrict the search space of a chart parser because only input which corresponds to the expectations is treated. The internal td-hypos built in an INITIAL PREDICT operation correspond to all available grammar rules with the specified properties for this operation.

The integration of external td-hypos in the parsing algorithm is only motivated under the following condition:

The intersection of the set of adapted td-hypos (TDH_{ext}) and the set of internal td-hypos (TDH_{int}) is a proper subset of TDH_{int}.

5.2 The Scanner

The problem of mapping signal time to chart nodes is solved within the SCAN operation in the modified ACP. In addition to the conventional linear precedence (LP) relation, incoming bu-hypos have a temporal dimension. A hypothesis not only contains a linguistic description of a subpart of the speech signal like it is in written language applications but also the interval of the speech signal which is associated with this linguistic description. The boundary points of this interval refer to signal time; the fuzziness of phonological events in relation to the signal, mainly due to co- articulation of phonological units, results in the problem of identifying and eliminating gaps and overlaps between hypotheses, known as *lattice-to-chart-mapping.*

In connection with non-incremental parsing, there exists a solution to this problem (1). However, no general solution has been found with respect to incremental chart analysis. In this section, a proposal for using a heuristic method with statistical support is made for this mapping; the modification of the scanner concerns the following modifications:

- Manual and automatic parametrisation of the (maximal) size of gaps and overlaps is permitted.
- A data structure for boundary points and related chart nodes sorted according to their temporal order is introduced for the mapping of new boundary points.

This treatment of the SCAN operation has consequences for the COMPLETE operation. The input for the morphological parser was defined as a lattice with competitive overlapping hypothesis. One consequence of this is that after the application of the lattice-to-chart mapping, two items which indeed fulfil the abstract connectedness condition defined by the COMPLETE operation, but are not temporally connected, are still possible.

For this reason, in addition to the manipulation of the boundary points in the lattice (the lattice solution), the SCAN operation is generalised so as to include a manipulation of the chart (the chart solution) as in (1). If an item $\mathrm{I}_{\mathrm{t\_inact}}$ has been added to the chart a special procedure for constructing functional edges I_{f} takes care of the connectedness of items which are not temporally connected but fulfil the connectedness condition. This condition states that two items are connected if the gap between them is not filled with another item whose beginning and ending points lie in this interval.

Instead of choosing both lattice and chart solutions for the lattice-to-chart problem another possibility is to take over the real boundary points into the COMPLETE operation and manipulate only the completor. This solution requires further quantitative information about the admissable completion over gaps and overlaps, too. This modification is not yet implemented.

5.3 The completor

The neccessary distinction between written language and speech leads to a distinction between a standard completor as used for written language (WL-COMPLETE) and a modified completor for spoken language (SL-COMPLETE) which can handle the manipulated chart. The complete operation is performed if an inactive item is added to the chart. Completion is iterative. All items in the chart which fulfil the complete premises are processed. The COMPLETE operations are defined as follows.

Let $G =< V_{\mathrm{N}}, V_{\mathrm{T}}, S, R >$ and input $w = a_{1}, ..., a_{\mathrm{n}}$

- WL-COMPLETE: For $0 \leq i \leq k \leq j \leq n$: if $A \rightarrow a.Bb \in chart[i,k]$ and $B \rightarrow y. \in chart[k,j]$, then $A \rightarrow aB.b \in chart[i,j]$.

- SL-COMPLETE: The premise of WL-COMPLETE or the following premise has to be fulfilled:
  For $0 \leq i \leq k1 \leq k2 \leq j \leq n$:
  if $A \rightarrow a.Bb \in chart[i,k1]$ and $B \rightarrow y. \in chart[k2,j]$ and $I_{\mathrm{f}} \in chart[k1,k2]$, then $A \rightarrow aB.b \in chart[i,j]$.

6 Conclusions

An interactive parametrised morphological chart parser for spoken language system development was described. The parser is designed for testing different parametrisations and interactions between morphological analysis and other components for the recognition of new words and corpus and speaker independent applications. Evaluation of the results is currently in progress.

References

[1] Lee-Feng Chien, K. Chen and Lin-Shan Lee. 1990. An Augmented Chart Data Structure with Efficient Word Lattice Parsing Scheme In Speech Recognition Applications. In *Proceedings of the 13th International Conference on Computational Linguistics*, pages 60–65, Helsinki, Suomi.

[2] Kenneth Church. 1987. *Phonological Parsing in Speech Recognition.* Kluwer Academic Publishers, Norwell, Massachusetts.

[3] Jochen Dörre. 1987. Weiterentwicklung des Earley-Algorithmus für kontextfreie und ID/LP-Grammatiken. *LILOG-Report*, 28, Universität Stuttgart.

[4] A. Jusek et al. 1994. Detektion unbekannter Wörter mit Hilfe phonotaktischer Modelle. In *Mustererkennung 94*, 16, DAGM-Symposium Wien, Springer Verlag, Berlin.

[5] Markus Kesseler, Gunther Görz. 1994. Anytime algorithms. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING)*, Kyoto, Japan.

[6] Kimmo Koskenniemi. 1983. *Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production*, Department Of The General Linguistics, University Of Helsinki, Helsinki. .

[7] Wolfgang Menzel. 1994. Parsing of Spoken Language under Time Constraints. *Proceedings of ECAI 94*, pages 560-564, August 1994, Amsterdam.

[8] Hermann Ney. 1984. The Use of a One-Stage Dynamic Programming Algorithm for Connected Word Recognition. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-32, No.2, pages 263-271, April 1984.

[9] Elmar Nöth,/ B. Plannerer. 1993. Schnittstellendefinition für den Worthypothesengraphen. *Verbmobil-Memo*, 2-94.

[10] Claudius Pyka. 1991. Deterministische, inkrementelle und zeitsynchrone Verarbeitung und die Architektur von ASL-Nord. *ASL-TR-22=91*, Universität Hamburg, Hamburg.

[11] Fernando C.N. Pereira, David H.D. Warren. 1983. Parsing As Deduction. In *Proceedings of the 21st ACL*, pages 137–144.

[12] Heike Rautenstrauch. 1994. Schritthaltende Generierung von Wortgraphen. In *DAGA 94*, pages 1261-1264, Dresden, Germany. DPG-GmbH.

[13] Susanne Riehemann. 1993. Word Formation in Lexical Type Hierarchies. A Case Study of bar-Adjectives in German. *SfS-Report-02-93*, Seminar für Sprachwissenschaft, Universität Tübingen, Tübingen.

[14] Graeme D. Ritchie et al.. 1992. *Computational Morphology.* MIT Press, London.

[15] Harald Trost. 1993. Coping with Derivation in a Morphological Component. In *Proceedings of the Sixth Conference of the European Chapter of the Association foer Computational Linguistics*, pages 368–376, Utrecht, The Netherlands. Association for Computational Linguistics, Morristown, New Jersey.