

Lecture Notes in Artificial Intelligence 1735

Subseries of Lecture Notes in Computer Science

Edited by J. G. Carbonell and J. Siekmann

Lecture Notes in Computer Science

Edited by G. Goos, J. Hartmanis and J. van Leeuwen

Springer

Berlin

Heidelberg

New York

Barcelona

Hong Kong

London

Milan

Paris

Singapore

Tokyo

Jan Willers Amtrup

Incremental Speech Translation



Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Author

Jan Willers Amtrup
New Mexico State University, Computing Research Laboratory
P.O. Box 30001, Las Cruces, NM 88003, USA
E-mail: jamtrup@crl.nmsu.edu

Cataloging-in-Publication data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Amtrup, Jan:

Incremental speech translation / Jan Willers Amtrup. - Berlin ;
Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ;
Paris ; Singapore ; Tokyo : Springer, 1999
(Lecture notes in computer science ; 1735 : Lecture notes in artificial
intelligence)
ISBN 3-540-66753-9

CR Subject Classification (1998): I.2.7, I.2, F.4.2-3

ISBN 3-540-66753-9 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1999
Printed in Germany

Typesetting: Camera-ready by author
SPIN: 10750071 06/3142 - 5 4 3 2 1 0 Printed on acid-free paper

Foreword

Human language capabilities are based on mental procedures that are closely linked to the time domain. Listening, understanding, and reacting, on the one hand, as well as planning, formulating, and speaking, on the other, are performed in a highly overlapping manner, thus allowing inter-human communication to proceed in a smooth and fluent way.

Although it happens to be the natural mode of human language interaction, incremental processing is still far from becoming a common feature of today's language technology. Instead, it will certainly remain one of the big challenges for research activities in the years to come. Usually considered difficult to a degree that renders it almost intractable for practical purposes, incremental language processing has recently been attracting a steadily growing interest in the spoken language processing community. Its notorious difficulty can be attributed mainly to two reasons:

- Due to the inaccessibility of the right context, global optimization criteria are no longer available. This loss must be compensated for by communicating larger search spaces between system components or by introducing appropriate repair mechanisms. In any case, the complexity of the task can easily grow by an order of magnitude or even more.
- Incrementality is an almost useless feature as long as it remains a local property of individual system components. The advantages of incremental processing can be effective only if all the components of a producer-consumer chain consistently adhere to the same pattern of temporal behavior. Particularly for inherently complex tasks like spoken language translation the issue of incremental processing cannot be treated as a local phenomenon. Instead it turns out to be intrinsically tied to fundamental questions of the overall system architecture, thus requiring a global perspective and the ability to create and maintain a sufficiently ambitious experimental environment.

If, despite these difficulties, a first prototypical solution for the incremental translation of spoken dialogues is presented here, two fundamental ideas have contributed most to this remarkable success: the use of a chart as a uniform data structure throughout the system and the rigorous application of results from graph theory that eventually allowed the complexity of the task to be reduced to a manageable degree.

This combination of contributions enables us for the first time to observe how a machine translation of natural language utterances evolves over time as more and more input becomes available. It certainly is much too early to risk a direct comparison with human interpretation capabilities, but certainly this book puts forward a benchmark against which other solutions will have to be measured in the future.

October 1999

Wolfgang Menzel

Preface

Automatic speech recognition and processing has received a lot of attention during the last decade. Prototypes for speech-to-speech translation are currently being developed that show first impressive results for this highly complex endeavor. They demonstrate that machines can actually be helpful in communicating information between persons speaking different languages. Simple tasks, e.g. the scheduling of business appointments or the reservation of hotel rooms and air travel tickets, are within reach.

Needless to say, the power of these prototypes is far from being equal to the human abilities for speaking, hearing, understanding, and translating. Performing the translation of speeches or free dialog at a high level is one of the most ambitious goals of scientists in the natural language processing domain. Several major areas of research have to be fruitfully combined to create even basic systems and demonstrators. Progress is needed regarding each of the several steps that are performed while creating a translation of an utterance spoken by a speaker, involving fields like acoustics, speech recognition and synthesis, prosody, syntactic processing, semantic representation, contrastive studies for translation, and many others.

This book starts from an outside view to speech translation, a view that does not concentrate immediately on one of the tasks we mentioned. The main motivation for the research presented in this monograph is the fact that humans understand and translate while they are still hearing. This *incremental* operation is in part responsible for the relative ease with which we handle certain tasks, like simultaneous interpreting or simply following a conversation at a party with a high level of background noise.

The application of this paradigm to automatic speech processing systems seems to be a natural thing to do, yet it has serious consequences for the implementation of individual components and the system as a whole. The simple demand “Start analyzing while the input is still incomplete” in some cases requires difficult modifications to the algorithms employed for certain tasks.

We think that incremental, modular systems require careful attention as to how they are composed from individual components. Interfaces and their use become more crucial if a component is to deliver not only a small set of final results (in many cases exactly one result), but a continuous stream of hypotheses. Thus, the realization of incrementality also initiated the use of an integrated data structure (the layered chart) and the use of a uniform formalism for all modules.

The first chapter introduces incrementality and provides a motivation for its use in automatic speech translation. Chapters 2 and 3 give the necessary theoretical foundation to describe the system presented here adequately. In particular, chapter 2 focuses on graph theory and its application to natural language processing. We believe that a wide range of phenomena and algorithms for NLP can be most adequately described (and most easily understood) in terms of a small subset of graph theory. Chapter 3 presents the uniform formalism that is used throughout the system: Typed Feature Structures.

Chapter 4 and 5 describe the system that provides the background for this book. In our opinion, interesting architectural paradigms cannot be shown in isolation from an actual system implementing these paradigms. The system MILC demonstrates the feasibility of employing incrementality to a complete speech translation system. We describe how the three architectonic principles incrementality, integrity, and uniformity are used to compose a non-trivial system, and demonstrate its performance using actual speech data. Finally, chapter 6 provides a conclusion.

The research described in this book was performed while the author was a research scientist in the Natural Language Systems Group within the Computer Science Department of the University of Hamburg, Germany. The German version of this monograph was accepted as dissertation by its CS department.

I am indebted to the teachers that provided a major part of my education in computer science. Walther von Hahn introduced me to Natural Language Processing. Günther Görz attracted me to syntactic parsing and the architecture aspects of NLP systems. Finally, Wolfgang Menzel discussed large parts of my work and contributed a lot as my primary thesis advisor.

The Natural Language Systems group in Hamburg provided an excellent research environment, from which I benefited during the five years that I worked there. I wish to thank all colleagues for their cooperation, especially Andreas Hauenstein, Henrik Heine, Susanne Jekat, Uwe Jost, Martin Schröder, and Volker Weber.

While in Hamburg, I worked as a Verbmobil project member for more than three years. This enabled me to have many fruitful discussions and to cooperate with several people, especially Jan Alexandersson, Thomas Bub, Guido Drexel, Walter Kasper, Marcus Kessler, Hans-Ulrich Krieger, Joachim Quantz, Birte Schmitz, Joachim Schwinn, Jörg Spilker, and Hans Weber.

Some of the work presented here (including most of the experiments) was finished at the Computing Research Laboratory at New Mexico State University. Special thanks to Karine Megerdooian, Sergei Nirenburg, and Rémi Zajac.

August 1999

Jan Willers Amtrup

Contents

Overview	1
1. Introduction	3
1.1 Incremental Natural Language Processing	3
1.2 Incremental Speech Understanding	11
1.3 Incremental Architectures and the Architecture of MILC	15
1.4 Summary	24
2. Graph Theory and Natural Language Processing	25
2.1 General Definitions	25
2.2 The Use of Word Graphs for Natural Language Processing Systems	30
2.3 Evaluation of Word Graphs: Size and Quality Measures	34
2.4 Evaluation of Word Graphs: Quality Measures	44
2.5 Further Operations on Word Graphs	48
2.5.1 Removing Isolated Silence	48
2.5.2 Removing Consecutive Silence	49
2.5.3 Removing All Silence Edges	51
2.5.4 Merging Mutually Unreachable Vertices	51
2.6 Hypergraphs	52
2.6.1 Formal Definition of Hypergraphs	54
2.6.2 Merging of Hyperedges	56
2.6.3 Combination of Hyperedges	59
2.7 Search in Graphs	60
2.8 Summary	62
3. Unification-Based Formalisms for Translation in Natural Language Processing	65
3.1 Unification-Based Formalisms for Natural Language Processing	65
3.1.1 Definition of Typed Feature Structures with Appropriateness	68
3.1.2 Type Lattices	68
3.1.3 Feature Structures	69
3.1.4 Functions as Values of Features	73
3.2 Unification-Based Machine Translation	73
3.3 Architecture and Implementation of the Formalism	76
3.3.1 Definition and Implementation of Type Lattices	79
3.3.2 Definition and Implementation of Feature Structures	80
3.4 Summary	84

4. MILC: Structure and Implementation	85
4.1 Layered Charts	86
4.2 Communication Within the Application	95
4.2.1 Communication Architecture of an Application	96
4.2.2 Channel Models	98
4.2.3 Information Service and Synchronization	100
4.2.4 Termination	104
4.3 Overview of the Architecture of MILC	105
4.4 Word Recognition	106
4.5 Idiom Processing	108
4.6 Parsing	110
4.6.1 Derivation of Verbal Complexes	111
4.6.2 Spontaneous Speech and Word Recognition	113
4.6.3 Structure and Processing Strategies	115
4.7 Utterance Integration	121
4.8 Transfer	128
4.8.1 Chart-Based Transfer	130
4.8.2 The Implementation of Transfer for MILC	132
4.9 Generation	137
4.10 Visualization	143
4.11 Extensions	145
4.11.1 Extension of the Architecture	147
4.11.2 Anytime Translation	149
4.12 System Size	152
4.13 Summary	152
5. Experiments and Results	155
5.1 Hypergraphs	156
5.2 Translation	158
5.2.1 Data Material	158
5.2.2 Linguistic Knowledge Sources	159
5.2.3 Experiments and System Parameters	161
5.2.4 Evaluation	162
5.2.5 Extensions	164
5.3 Comparison With Non-incremental Methods	165
5.4 Summary	167
6. Conclusion and Outlook	169
Bibliography	175
Glossary	193
Index	195

List of Figures

1.1	The architecture of a functional unit within a modular system	4
1.2	The architectonic overview of MILC	17
1.3	Interlingua and transfer	19
1.4	Multi-level transfer	21
2.1	A chart for <i>Der Peter singt mit Freude</i>	29
2.2	The graph $K_{(3,3)}$	29
2.3	Word graph for the utterance n002k000	33
2.4	Unique word graph for the utterance n002k000	35
2.5	A difficult graph w.r.t the reduction to unique label sequences	40
2.6	Runtime for reducing graphs to unique label sequences	42
2.7	A complex graph for rank computation	46
2.8	Merging of silence edges	49
2.9	Two families of edges in a word graph	52
2.10	An interval graph	53
2.11	Two families of word hypotheses as hyperedges	54
2.12	Adding a word hypothesis to a hyperedge	58
2.13	Creation of additional paths by using hypergraphs	60
3.1	Feature structure for a simple syntactic rule	68
3.2	A transfer rule in LFG style	74
3.3	A transfer rule in TFS style	75
3.4	A transfer rule by Beskow	75
3.5	A small part of a type lattice	79
3.6	One part of the structure of lexical items	79
3.7	Feature structure for a simple syntactic rule	81
3.8	Definition of vertices within feature structures	82
4.1	The <i>Whiteboard</i> -Architecture	89
4.2	The principle layout of a layered chart	91
4.3	Principle component layout	97
4.4	The configuration of <i>split channels</i>	99
4.5	An example of a configuration file for split channels	101
4.6	Time line of the initial configuration of channels with the ILS	103
4.7	XPVM snapshot of the initial synchronization	104

4.8	The architectonic overview of MILC	106
4.9	A word graph in Verbmobil syntax	107
4.10	Idiom definition for the partial utterance “tut mir leid” (“I am sorry”) ...	109
4.11	A syntactic rule for verbs in last position with four complements	112
4.12	One of the lexical entries for “Arbeitsstreffen” (“work meeting”)	116
4.13	A grammar rule for noun phrases with a determiner	118
4.14	A noun phrase from the partial parser	120
4.15	Island analysis: An example rule for nominal phrases	122
4.16	A rule for complements to the left of the verb	124
4.17	Lexicon entry for “ausmachen” (“schedule”)	126
4.18	A verbal phrase from the integrator	127
4.19	A transfer chart for structural translations	132
4.20	A transfer rule for verbal phrases	133
4.21	The transfer lexicon entry for “recht sein” (“suit”)	134
4.22	Topology of the fundamental rule for transfer	135
4.23	A subpart of the generation input for “lassen Sie uns das nächste Arbeitsstreffen vereinbaren” (“let us schedule the next work meet- ing”)	139
4.24	Generation rules for imperative verbs	140
4.25	Generation lexicon entry for “let”	142
4.26	A snapshot of the processing with MILC	145
5.1	Reduction of word edges by hypergraph conversion	156
5.2	Reduction of chart edges by hypergraph conversion	157
5.3	Reduction of analysis time by hypergraph conversion	158
5.4	Comparison of incremental and non-incremental processing	167

List of Algorithms

1	Computing the topological order of a DAG	36
2	Computation of the transcript independent density of a word graph	37
3	Calculation of the number of paths in a graph	38
4	Reducing a graph to unique label sequences	39
5	Merging of two vertices	41
6	Calculation of the number of derivation steps of a fictitious parser for a word graph	43
7	Computing the rank of a path (part 1)	44
8	Computing the rank of a path (part 2)	46
9	Removing isolated silence edges	49
10	Removing consecutive silence	50
11	Removing all silence edges.....	51
12	Adding a word hypothesis e_n to a hypergraph $G = (\mathcal{V}, \mathcal{E}, \mathcal{L}, \mathcal{W})$	58
13	SSSP for DAGs	61
14	SSSP for incremental hypergraphs.....	62

List of Tables

3.1	The syntax for type lattices	80
3.2	The syntax of feature structures	82
3.3	Internal representation of a feature structure	83
4.1	Messages sent from the word recognizer to the successor components . . .	108
4.2	Results for standard and partial parsing of the dialog n002k	115
4.3	Sketch of the processing of “lassen Sie uns den nächsten Termin ausmachen” (“let us schedule the next appointment”)	125
4.4	An example for the output of the generator	143
4.5	System size (in lines of code)	153
5.1	Properties of the dialogs used in the experiments	159
5.2	The utterances in dialog m123	160
5.3	Results of analyzing five dialogs	162
5.4	Generator output for the utterance j534a005	163
5.5	Evaluation of the translations	164
5.6	Comparison of runtime for incremental and non-incremental configurations	165

Overview

Human language understanding works incrementally. This means that people process parts of the acoustic input, even before it has been completed. Engaged in a dialog, they are able to anticipate their partner's contribution before she or he has stopped talking. Simultaneous interpreters provide an example relevant for the purpose of the research presented here: They start to produce target language utterances with little delay, thereby showing incremental behavior.

Consequently, the introduction of incremental strategies into artificial systems for understanding natural language seems reasonable to implement adequate processing which exhibits at least part of the human performance. If machines are expected to perform somewhat similarly to humans, incremental algorithms have to be incorporated.

There is still some hesitation, however, to investigate incremental mechanisms. This is mainly due to the fact that incorporating these strategies results in an increased processing burden in the first place, because there is less certainty about the right context. This increase in processing time can be partly countered by employing means of parallelization which become possible in incremental systems. A gain in quality of individual processing steps can only be reached by constructing alternative information paths used for mutual influence between two components.

The predominant area in which investigation into incremental systems is pursued nowadays seems to be the study of speech-to-speech translation. First promising results have already been achieved.

The system MILC (*Machine Interpreting with Layered Charts*), which is developed and described in this work, constitutes the attempt to design and implement an application for translation of spontaneous speech which works incrementally throughout. Each component of the system starts operating before preceding components have finished their processing. Parts of the input are worked on in strict order of speaking time. This schema results in an incremental behavior of the application as a whole, which is able to generate target language text before the source language speaker has finished his or her turn.

In order to realize such a system in an integrated and uniform way, we develop a new data structure, called *Layered Chart*, which is used to store temporary results throughout the system. This structure extends the way in which charts have been used so far, namely for parsing and generation. Additionally, we implement a typed feature formalism which aids in maintaining the knowledge sources for the

system. This formalism is specifically designed to simplify the means of communicating linguistic objects between different modules of a system. Both measures linked to each other guarantee a homogeneous view of the global state of the translation system and enable the investigation of non-trivial interaction patterns between components.

The successful implementation proves that it is possible to realize an integrated system for interpreting spontaneous speech for restricted corpora in an architecturally strict way. Moreover, MILC provides the basis for more far-reaching investigations into the architecture of incremental speech-processing systems. This is exemplified by the incorporation of the incremental recognition and translation of idioms.

This monograph is organized as follows:

- Chapter 1 provides a detailed introduction to the central matters of this work. We present the incremental nature of human speech comprehension and show how incrementality is used within natural language processing.
- Chapter 2 contains an overview over the relevant parts of graph theory which have an impact on speech processing. We present some results from the evaluation of speech recognition. Additionally, we introduce hypergraphs which are a basic data structure for the representation of highly redundant recognition results.
- Chapter 3 gives an introduction to unification-based formalisms and describes their use in NLP systems with an emphasis on machine translation. We present the typed-feature formalism which was implemented for the system described here.
- Chapter 4 describes the architecture and implementation of the MILC system. We focus on the global architecture, the communication subsystem, and some properties of individual components.
- Chapter 5 contains the description of the data and experiments we used to evaluate MILC. The different knowledge sources, which are relevant for the evaluation, are presented.
- Chapter 6 enumerates the central results presented throughout this book and tries to give an outlook on future research in the field.