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

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 Kesseler, 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 Megerdoomian, Sergei Nirenburg, and Rémi Zajac.

August 1999 Jan Willers Amtrup

Contents

List of Figures

List of Algorithms

List of Tables

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 (*M*achine *I*nterpreting with *L*ayered *C*harts), 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

J.W. Amtrup: Incremental Speech Translation, LNAI 1735, pp. 1-2, 1999.

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.