

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.

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 Megerdooimian, Sergei Nirenburg, and Rémi Zajac.

August 1999

Jan Willers Amtrup