Architecture as a Problem of Information Fusion

Wolfgang Menzel
Universität Hamburg, Fachbereich Informatik

Abstract

In order to successfully integrate the contributions of different components in a complex system for spoken language processing, powerful and flexible techniques for information fusion are required. If applied at an appropriate level of granularity and supporting the functional autonomy of individual components based on a preferential combination of individual contributions, such an approach should allow to combine the available evidence from different sensory channels or independent knowledge sources and therefore might provide for higher accuracy and robustness. Experiences with the application of information fusion techniques within the Verbmobil system are reviewed and the prospects of an architecture based on weighted constraint dependency grammars are discussed.

Spoken language processing requires to combine very different evidence from a wide variety of knowledge sources. This includes phone characteristics, prosodic cues, the lexical inventory, possibilities for combining lexical units into larger syntactic constructions, as well as their semantic coherence and pragmatic plausibility. Language processing in general can then be understood as the problem of bringing together all the relevant knowledge pieces for the task at hand and eventually deciding about the probably best target representation for an incoming utterance or a given communicative goal.

From a decision theoretic point of view, the optimimum performance can only be expected if all the available evidence is combined into a single objective function. This suggests a monolithic system architecture where all the different aspects of language are comprised in a simple and uniform model structure. This view on the problem of system architecture has its roots in the pattern recognition tradition of spoken language processing, where the task of utterance recognition is decomposed into just two different components, namely the acoustic model, describing the probability of the observed signal given a particular word hypothesis, and the language model approximating word sequence probabilities. Nevertheless, the contributions of both components are fed into a single decision procedure, which tries to determine the optimum hypothesis given a speech signal observation. The only reason for separating the two components at all has to do with the very practical consideration that the acoustic and the language model can and should be trained on quite different kinds of data.

More recently, the same principles of system decomposition have been applied to other language processing tasks like syntactic parsing or machine translation. In all those cases, a very small number of model components, which capture the input-output mapping, is identified and integrated by means of a single decision rule based on Bayes' theorem. For machine translation, for instance, the relation between a source language sentence and its possible counterparts in the target language is described by the language model of the target language and a string translation model, which in turn is broken down into lexical correspondences and source-target-string alignments.

This approach, although ideal from the decision theoretic point, comes with several disadvantages. Most obvious is the opacity of such a monolithic kind of model. Being trained to support relatively complex decision procedures, it usually comprises information which belongs to a wide variety of different phenomena, merged in a completely unseparable fashion. Acoustic phone models, for instance, are neither able to distinguish between prosodic nor dialectal variants. If they are designed to do, this results in considerably more phone classes which, of course, are more difficult to train. Language models, on the other hand, comprise syntactic, semantic and pragmatic evidence, which also cannot be separated easily. This is particularly problematic from the point of view of portability, where we are faced with a kind of complementary behaviour: Usually, syntactic regularities easily carry over to new domains, but large parts of semantics do not. On the other hand, porting to other languages happens to be straightforward for large parts of semantics, while it is almost impossible for syntax.

Moreover, automatic training procedures are applicable only as long as sufficient training data can be supplied. Thus, they lend themselves for the more general subtasks, like part of speech tagging, whereas application specific aspects seem to require more traditional customisation possibilities. During training a clear factorization between general and task specific aspects becomes necessary, which is difficult to achieve in monolithic architectures. For the same reason, monolithic approaches render any attempt to integrate additional knowledge after model completion almost infeasible. This makes them particularly unattractive for interactive solutions to language processing.

From a knowledge engineering point of view a more diversified approach seems to be preferable, where the system is decomposed into a much larger set of fairly independent submodules, which then can be developed, evaluated and maintained according to their specific requirements. This has been the predominant methodology of symbolic language processing, which is guided by linguistic insights into language and its different description layers. The customary distinctions between phonetics and phonology, syntax, semantics and pragmatics etc. suggest a *horizontal decomposition* of the architecture into functional components.

The need for horizontal decomposition is also supported by observations about the astonishing flexibility of the human language processing system. It not only facilitates the mutual compensation of deficits based on partial complementarity and structural redundancy, but in case of communicative problems also allows the hearer to clearly distinguish between different reasons for failure: *This sounds strange* vs. *This is ungrammatical* vs. *This is bad style* vs. *This is nonsense* vs. *This is not true* etc. It also allows human speakers to reuse significant *parts* of their language system in quite new communicative settings, e.g. by relying on fairly general syntactic knowledge when aquiring the semantics of a language for special purpose or by reusing most of native language semantics when learning a foreign one. Neither of these settings requires to start language learning all over again or necessarily involves the loss of already existing language capabilities.

In order to facilitate the independence of system components, modules are usually designed to carry out independent decision procedures based on the locally available information. Such decisions, of course, are necessarily suboptimal from the global perspective of system performance. To cope with the nonlocal consequences of misleading local decisions turns out to be one of the most serious problems in the design of complex language processing systems. In general, an architecture which communicates an array of weighted alternative solution candidates (word hypotheses graphs (Oerder and Ney, 1993), constituent charts (Amtrup, 1999), etc.) has turned out to be superior to making deterministic local decisions, since it avoids the necessity to correct wrong local decisions later.

There is, however, another dimension along which a system can be broken down into independent processing tasks. This possibility is given whenever there are different but complete decision procedures available which all are able to produce the desired results, but do so with a complementary decision quality (or sometimes fail to deliver something at all). In general, such a *vertical* decomposition facilitates a kind of competitive approach and gives rise to the hope that quality of the overall arrangement can be improved beyond the contributions of the individual components. It requires, however, a suitable selection component, which can be used to determine the probably best solution from an array of alternative candidates.

Another situation in which vertical decomposition seems to be useful is the exploitation of suprasegmental information in different processing components of a complex language processing system. Here, the different modules are certainly not meant to derive the same kind of processing result using different computational approaches, but rather to provide fully complementory information which eventually has to be combined in order to achieve the desired goal.

Both kinds of decomposition give rise to the need for information fusion in a complex system architecture. While horizontal decomposition in combination with the communication of decision alternatives more or less amounts to combining the different models into a monolithic one (perhaps with a few concessions regarding completeness of the underlying search), information fusion in a vertically decomposed system poses completely new questions. Most obvious is the question about the appropriate granularity level on which the different contributions can be brought together sucessfully. Equally difficult seems to be the appropriate definition of a comparison measure needed to evaluate the usefulness of the available information pieces. It is this second type of information fusion problem I will focus on here in this paper.

1 Information fusion

A number of quite different reasons can be identified which speak in favor of information fusion techniques when designing a complex system architecture:

1. the necessity to combine information from different sensory channels like in lipreading or automatic subtitle synchronisation. Here, the need for information fusion is inherited from the processing task itself and can be avoided by no means. For lipreading, evidence from the acoustic and the visual stimulus has to be integrated, whereas subtitle synchronisation has to combine the speech signal extracted from the sound channel with the language data as given by the script. Horizontal decomposition is not available for this problem class.

Such a multi-channel situation, which is typical for all kinds of multi-modal information processing, obviously can be considered as the canonical case of information fusion. However, even in a single-modality system it might become necessary to integrate complementary information, which e.g. comes

2. from different data sources, which for practical or fundamental reasons cannot be acquired (i.e. learned) together. In a horizontally decomposed system such a situation is given for instance with the division of labour between acoustic and language model in speech recognition. Here, corpora of considerably different size are required to train the respective models. Vertical decomposition, on the other hand, is required for the treatment of segmental and suprasegmental information in spoken language processing. In both cases the desired information is derived directly from the speech signal but needs to be combined somehow further on in the processing chain. In principle, a monolithic approach should be possible, but seems almost infeasible.

3. from different processing components, which have been developed based on completely different paradigms, e.g. the combination of corpus-oriented machine learning approaches with others based on more traditional manual modelling techniques. If successful, it would allow to better interface automatically acquired models with existing knowledge sources like semantic databases. Another application is the integration of direct instruction techniques, which are needed e.g. in interactive environments.

4. from different components, which have been designed to separate different knowledge sources in order to improve on certain aspects of system behaviour like portability to other domains and languages or perspicuity with respect to system failure. This holds for instance for the relationship between syntax and semantics, which usually are decomposed horizontally but where a vertical approach seems more promising, since both can make complementary contributions to the construction of a logical form.

In the first two settings information fusion is used to provide the system with additional disambiguating evidence. The approach is based on the assumption that this additional information (like lip shape patterns or prosodic cues) might eventually increase the reliability of processing results. In the latter two cases structural redundancy is introduced, which might contribute to a higher degree of robustness. If, for instance, one or several components fail to deliver a result at all, possibly a fall-back solution is available. Which result to choose from a set of available hypotheses requires a voting component, which is able to judge different contributions with respect to their utility. In a more sophisticated arrangement the final result can also be created by combining different partial results coming from different components.

All these approaches are based on the assumption that different components can provide somehow complementory information or achieve qualitatively different results for different problem instances. Therefore, an arrangement based on the idea of competition should be able to attain a higher degree of robustness if compared to its individual components.

To expect such benefits, however, is only realistic as long as the chosen system architecture fulfills three important preconditions:

- Information fusion has to be attempted at a proper level of granularity, which allows to effectively support good local choices. Prosodic cues, for instance, can provide important contributions to a wide variety of decisions within a complex language processing system, but do so on very different levels, namely lexical disambiguation, the segmentation of utterances, the structural attachment of discourse particles, or the assignment of pragmatic functions, to name a few.

- An assessment function is required to evaluate the different information pieces according to their potential contribution to the processing goals. This function can be based on a self-evaluation of the respective processing component (usually given as a kind of confidence score), or an external evaluation based either on purely formal criteria (like the spanning length of a partial solution), or a complex model of linguistic competence. The availability of such an assessment function is most crucial if alternative (partial) solutions have to be integrated into a single one. From a theoretical point of view the approach based on a linguistic competence model seems most attractive. It is, however, also faced with the most severe practical difficulties. There is a strong, principled argument against such an attempt: If a component for the evaluation of and the selection among alternative hypotheses already has the necessary information to do the optimal choice, why not use it to do the whole processing task on its own?

- The mutual dependence of components should be minimized, in order to avoid system failures due to missing or conflicting information. This requires a flexible integration scheme based on some kind of soft-computing approach like probability theory or fuzzy-sets. If suitably used it might prevent a failing component from turning down the whole processing chain.

2 Vertical information fusion in Verbmobil

Verbmobil as a system for speech-to-speech translation can be seen as a prototypical example of a highly complex system for spoken language processing. It is decomposed into a huge number of modules (cf. Figure 1) which are organised in several alternative translation paths, among them an example-based approach, a stochastic translation, a shallow translation based on dialog act information, and a transfer-based deep translation receiving its input from a range of different analysis modules.

Within the Verbmobil-system the need for information fusion arises in quite different places. Since only one sensory channel (the microphone) is used, there is no need to combine alternative ones. All three other types of information fusion as discussed above actually can be found in the final demonstrator. I am going to discuss some of these in the subsequent sections.

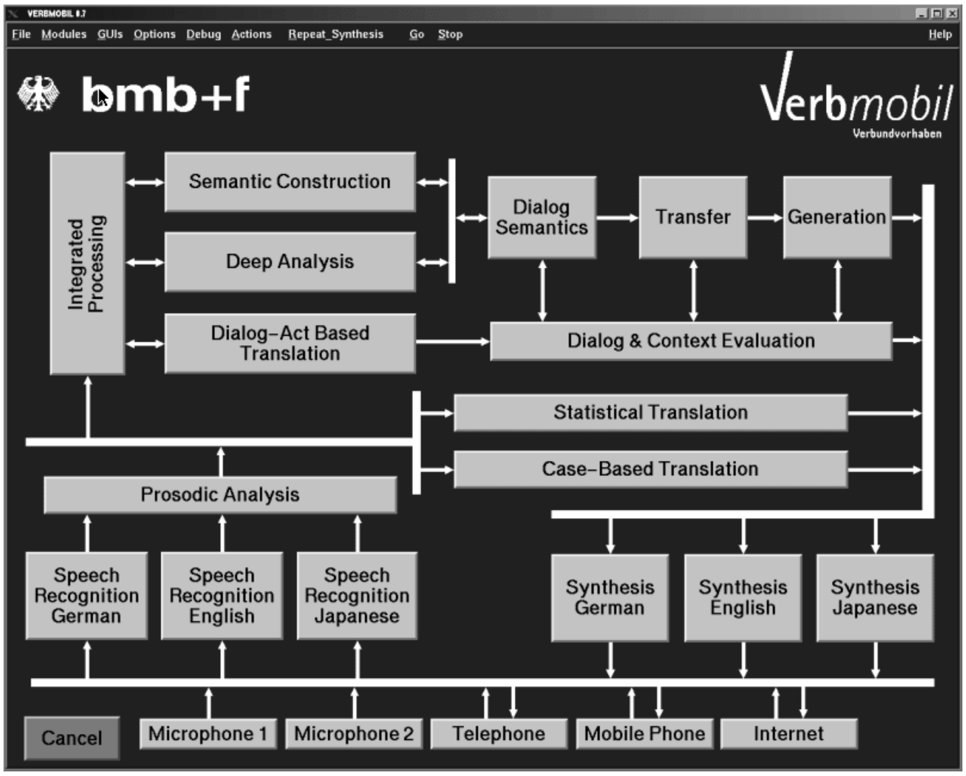

Figure 1: The control panel of Verbmobil.

2.1 Integrating evidence from different sources

Combining suprasegmental categories like accent information, intonation patterns or phrase boundaries smoothly with their segmental counterparts is one of the great challenges in the design of complex architectures for spoken language processing. Within Verbmobil this integration has been successfully accomplished in several places

utterance segmentation: A prosodic event detector coupled with a language model is used to segment dialog turns into shorter ”utterances”, which then are classified into dialogue acts (Batliner et al., 2000). This, in fact, is a horizontal decomposition since information is not really combined but simply propagated to a subsequent module in a sequential architecture. No information fusion in the proper sense takes place.

priority ranking of partial hypotheses: Within the HPSG-based ”deep” analysis component of Verbmobil (Kiefer, Krieger, and Nederhof, 2000) partial results are ordered according to being consistent with predictions of the prosody module about sentence and inter-sentential boundaries. Chart entries are penalized when crossing a predicted boundary or promoted when their endpoint coincides with a mandatory boundary. In both cases the ranking is influenced by the likelihood of the boundary. This priority ranking does not rule out any partial reading licensed by the grammar, but rather tries to delay unlikely alternatives.

Information fusion is preferential since hand-coded weights from the lexicon and the grammar are combined with the prosodic information and an additional spanning length criterion. Moreover, functional independence is granted, since the prosodic information is not vital for the normal working mode of the parser itself.

A similar approach has already been adopted by Kompe et al. (1997), where a substantial reduction in ambiguity (96%) and parsing time (92%) was achieved with only a minor drop in the success rate (2.5%).

semantic disambiguation: As a kind of fall-back position prosodic information about sentence mood is used to distinguish questions from non-questions, whenever a set of heuristic rules happens to fail on this task. Similarly, accent information is used to disambiguate the translation of focus particles (Bos and Heine, 2000). In both cases information fusion neither is preferential (likelihood estimations are not taken into consideration) nor provides functional independence.

supporting the stochastic translation: The stochastic translation component of Verbmobil uses prosodic boundaries and sentence mood information to insert delimiters into the recognised word form sequence. These segment boundaries are then used to restrict the space of possible alignments (Vogel et al., 2000). Integration is non-preferential but functional independence is given.

Undoubtly, the most ambitious integration scheme has been implemented with the preference ranking approach of the deep analysis. Unfortunately, for none of the integration efforts evaluation results are available, which would allow to estimate the real contribution of prosodic cues to the overall performance of the system.

2.2 Coupling of alternative processing components

The most obvious case of a competition-based vertical decomposition in Verbmobil is the availability of several alternative translation paths (multi-engine approach Wahlster (2000)). It creates the need to include a selection component, which has to evaluate the available translation alternatives to choose the most appropriate one (Cavar, Küssner, and Tidhar, 2000). The selection is based on

- a self-assessment of the contributing components by means of confidence scores which are rescaled in an iterative training procedure,
- dialog act information, which takes care of the fact that translation quality for some components highly depends on the dialogue act under consideration,
- the observation of certain ambiguity types, which are more reliably resolved by the deep processing components.

Unfortunately, contradictory results have been reported for the quality of the selection. In a first evaluation the component was able to improve the number of good translations by 27.8% compared to the quality level of the best component available at that time. However, due to changed parameters in the end-to-end evaluation scenario, these results were not replicable in a later evaluation effort (Tessiore and v. Hahn, 2000), where the percentage of "approximately correct" turn translations dropped 17.4% below the best component, a result which can be improved again using a trained selection procedure raising the overall quality level to 13.0% above the best individual component.

One might speculate about possible reasons for the partially unsatisfying results. Obviously, in certain cases the selection component somehow tends to "level out" the overall quality instead of reliably determining the best translation. It seems that information fusion for complete translations does not provide the optimal level of granularity. At that late a stage in the overall processing little information is available to determine why one translation might be better than another one. Perhaps this would suggest to attempt an earlier fusion of intermediate and partial results, thus achieving a much higher degree of system integration.

Another good example for a competion-based vertical decomposition is the integration of partial results from different parsing approaches for the syntactic-semantic analysis (Block and Ruland, 2000) which are combined by a component for robust semantic interpretation (Rupp et al., 2000). The approach is based on a common interface specification, namely sequences of

Verbmobil interface terms (VITs). Its main advantage is the possibility to combine several partial hypotheses in those cases where no spanning interpretation of the whole utterance has been produced by any component.

If, however, alternative (partial) hypotheses are available, the selection component not only decides between the alternatives, but tries to find an optimal combination of partial and fragmentary analyses into larger VITs. This combination is accomplished in a left-to-right incremental manner, i.e. under the condition of incomplete information. The combination of two subsequent VITs is blocked whenever both have been separated by a prosodic boundary (Rupp et al., 2000).

The decision on the optimal combination is based on three (formal) selection preferences

1. parser results are better than those obtained by one of the shallow components
2. integrated results are better than fragmented ones
3. longer analyses are better than short ones

No quantitative evaluation of this selection mechanism has been published. An attempt to train the selection model using more content-oriented selection criteria for VITs failed, since it obviously turned out quite difficult to generalize complex VITs to abstract descriptors preserving the crucial information needed for making well informed decisions.

2.3 Deliberate separation of model components

There seems to be only one example for a deliberate modularization in order to factor out different informational contributions. This is the syntax-semantics distinction of HPSG where syntactic constraints (i.e. subcategorization frames with their case restrictions) are separated from semantic constraints (i.e. functor-argument structures with selectional restrictions). However, these different contributions are coupled conjunctively thus abandoning the ideas of preferential combination and functional independence.

3 WCDG as a framework for information fusion

Weighted constraint dependency grammar (Menzel, 1998; Heinecke et al., 1998; Schröder et al., 2000) is a formalism for the disambiguation of dependency structures. Constraints license partial structures (dependency edges or combinations thereof). They come with weights attached to them, which sometimes can even be computed dynamically depending on the structural configuration under consideration. This possibility is used to model e.g. distance phenomena or dynamically changing preferences. In general constraints can be violated if no better structural description is available. The parser tries to determine the optimal structural interpretation according to an accumulatory function, which combines all the weights of violated constraints into an overall score for complex dependency trees.

The approach supports information fusion for two important reasons:

1. It invites to model different linguistic phenomena like syntax or semantics on separate description levels, which nevertheless are integrated into the same global optimization procedure. Thus, modularization is supported with all its advantages with respect to a higher degree of robustness and a better perspicuity of system behaviour, while local and therefore suboptimal decisions are avoided.
2. Since dynamically computable weights are available anyhow, constraints are an ideal interface to include a variety of additional preferential evidence like acoustic scores, prosodic cues, as tag probabilities or chunk boundaries into the global decision procedure.

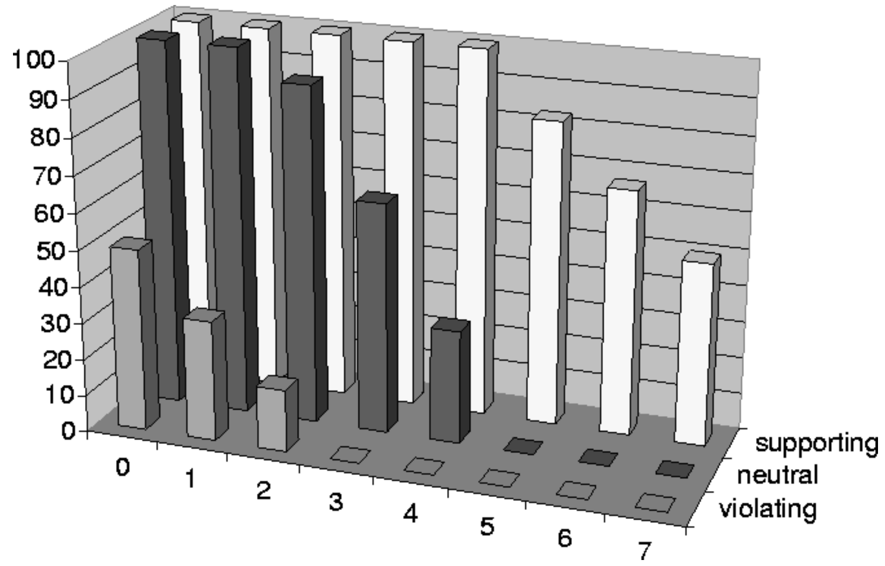

Figure 2: Mutual compensation of deficits between syntax and semantics. Sentence accuracy is plotted against severity of syntactic distortion, and against supportive or conflicting semantic and domain knowledge.

Due to its flexibility the approach facilitates experimentation with different architectural options and allows to rather precisely determine the individual contribution of specific information pieces to the quality of the processing results.

So far, experiments have been carried out to investigate

- the interplay between syntax and semantics
- the integration of acoustic scores from a speech recognizer, and
- the integration of probabilistic scores from a part-of-speech tagger.

3.1 Coupling syntax and semantics

The preferential coupling of highly autonomous components for syntactic and semantic disambiguation has first been tested on a relatively small domain from a language learning scenario (Menzel and Schröder, 1999). Here syntactic and semantic structures have been disambiguated on separate levels of description, each establishing its own dependency structure. These structures, however, are mapped onto each other by preferential constraints, like ”The subject tends to be the agent of the main verb.” Due to the partial redundancy introduced between the syntactic and semantic layer, a degree of robustness was established, which seems necessary to analyse and diagnose the highly ill-formed input typical for language learners without providing for specific error case treatments. Figure 2 shows the sentence accuracy for different degrees of syntactic and semantic distortion. Even in case of comparatively severe syntactic errors the intended semantics was obtained if semantic support (e.g. by means of selectional restrictions) was strong enough. On the other hand, strange semantic interpretations are identified relatively reliable if enough syntactic evidence (e.g. case government and various agreement conditions) is available.

More recently, the experiment has been repeated with another grammar for a small corpus of Verbmobil utterances. It turned out however, that the grammar was designed in a way which did not support functional independence sufficiently. If the semantic level was completely switched off, only a minor decrease of the performance on the syntactic structures was observed. Thus, in this particular case semantics did contribute almost nothing to syntactic disambiguation but rather introduced a new source of uncertainty. Nevertheless, the experiment confirmed the claim that due to the adopted modular approach, a precise evaluation of the different knowledge contributions to the overall performance of the system became possible.

3.2 Integrating a part-of-speech tagger

Usually, taggers are interfaced with a subsequent parser by having the tagger determine a unique tag sequence on which the parsing itself is carried out. For stochastic parsers the possible gain of letting the parser itself choose between alternative tags (multi-tagging) has turned out to not justify the additional computational effort (Charniak et al., 1996). In order to replicate this result for a constraint-based parser we interfaced the tagger by means of a single local constraint which penalizes all disagreements between the category of the corresponding lexical reading and the tag as proposed by the tagger. The penalty factor then is determined proportionally to the probability of the tag. Due to this preferential integration of tagging results the syntactic edge accuracy of the parser improved from 98.0% to 98.5% for the multi-tag approach, but decreased to 97.8% when only the single best tag was considered (Schröder, 2002).

This last result does not confirm the findings of Charniak et al. (1996). It suggests the conclusion that by relying on sequential patterns a Hidden Markov Model provides a complementary kind of information as compared to the contribution of dependency constraints, which have been defined on structural configurations with little attention to precedence regularities. Obviously, this complementarity is lost if both components are designed around a quite similar modeling paradigm.

3.3 Parsing with acoustic scores

The WCDG parser has been interfaced with a speech recognition engine using customary word hypotheses lattices as produced by the Verbmobil system (Schröder, 2002). Acoustic scores are re-normalized by means of a logarithmic mapping function and integrated into the grammar using a local constraint penalizing the incorporation of low-scored word hypotheses into the final structural interpretation of an utterance.

Two main results have been obtained:

- Due to the guiding information of acoustic scores, direct word lattice parsing became possible for graphs with a maximum density of 18.
- The restricting power of the grammar was not sufficient enough to allow to re-rank alternative recognition hypotheses in cases where the desired one did not receive the best acoustic score. Using the acoustically best word sequence was always superior to additionally considering alternative hypotheses.

The last result was mainly due to the fact that the grammar has not been developed with acoustically ambiguous input in mind. So it was never meant to be able to decide between two acoustically similar word strings like ″wir″ (we) or ″sie″ (you).

3.4 Competitive parallel computation

Using guided local search (Voudouris, 1997), a transformation-based optimization procedure with very attractive anytime properties, a coarse-grained parallelization of the parser has been attempted (Schulz, 2000). Two or more differently parametrized instances of the solution procedure are executed in parallel and notify themselves regularly about the best solution available so far. Due to this information exchange an increase in edge accuracy from 99.6% for the best individual parser to 99.8% for an arrangement with two components was achieved. Furthermore, the processing time remained well below the accumulated time required by the components running in isolation.

Obviously, in this case the coupling leads to a real synergy between the alternative processing paths. However, this arrangement profits considerably from using the same metrical space in both cases, which makes the comparison of alternative results very easy. On the other hand, the components differ only in the sequential order in which they traverse the problem space. Thus, comparatively little complementary information is available, since both make use of the very same grammar.

3.5 Further investigations

Other information fusion problems which can be investigated within the WCDG framework are the integration of prosodic information into the parsing procedure and the use of chunk boundaries. The idea is similar in both cases: Certain dependency edges crossing a boundary hypothesis for prosodic phrases or syntactic chunks are penalised. In contrast to the different attempts of integrating prosodic information in Verbmobil this approach does not consider boundary hypotheses as ultimate but utilizes the available confidence scores to make them defeasible as well. Thus, wrong decisions in the segmentation procedure can be overwritten by strong enough evidence from the language system. Furthermore, chunking results are integrated as early as possible into the decision procedure of a deep parser. Therefore this approach seems more promising compared to a later integration based on fully determined syntactic structures.

4 Conclusions

A vertical decomposition of a complex processing systems comes with a number of advantages, most notable with respect to robustness, flexible development and maintenance of components, as well as the combination of knowledge from different sources, but requires powerful techniques to bring the different contributions together again. An analysis of the solutions within Verbmobil has shown that existing approaches for information fusion in a vertically decomposed architecture still fall short of their potential. As long as the apparent shortcomings have not been overcome, the question of suitable architectures for language processing still needs to be considered an open one.

The ultimate goal will be to combine the advantages of monolithic and modular approaches: near-optimal decisions on the one hand but factorization of information contributions on the other. This would allow to effectively make use of achitectural features like structural redundancy, the combination of complementary contributions and a competition between alternative approaches. Weighted constraint dependency grammars can be seen as an important contribution on the way towards this goal, since they support integrated structural disambiguation in a strictly modular system design, provide for an arbitration of preferential evidence during the parsing process and therefore facilitate a flexible experimentation with different options for information fusion in order to eventually decide upon an optimal arrangement.

References

Amtrup, Jan W. 1999. *Incremental Speech Translation*. Lecture Notes in Computer Science, volume 1735. Berlin, Heidelberg, New York: Springer-Verlag.

Batliner, Anton, Jan Buchkow, Heinrich Niemann, Elmar Nöth, and Volker Warnke. 2000. The prosody module. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag, Berlin, Heidelberg, pages 106–121.

Block, Hans Ulrich and Tobias Ruland. 2000. Integrated shallow linguistic processing. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag, Berlin, Heidelberg, pages 143–146.

Bos, Johan and Julia Heine. 2000. Discourse and dialogue semantics for translation. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag, Berlin, Heidelberg, pages 336–347.

Cavar, Damir, Uwe Küssner, and Dan Tidhar. 2000. From off-line evaluation to on-line selection. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag, Berlin, Heidelberg, pages 597–610.

Charniak, Eugene, Glenn Carroll, John Adcock, Anthony R. Cassandra, Yoshihiko Gotoh, Jeremy Katz, Michael L. Littman, and John McCann. 1996. Taggers for parsers. *Artificial Intelligence*, 85(1-2):45–57.

Heinecke, Johannes, Jürgen Kunze, Wolfgang Menzel, and Ingo Schröder. 1998. Eliminative parsing with graded constraints. In *Proceedings 17th International Conference on Computational Linguistics, 36th Annual Meeting of the ACL, Coling-ACL '98*, pages 526–530, Montreal, Canada.

Kiefer, Bernd, Hans-Ulrich Krieger, and Mark-Jan Nederhof. 2000. Efficient and robust parsing of word hypotheses graphs. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag, Berlin, Heidelberg, pages 280–295.

Kompe, R., A. Kießling, H. Niemann, E. Nöth, A. Batliner, S. Schachtl, T. Ruland, and H. U. Block. 1997. Improving parsing of spontaneous speech with the help of prosodic boundaries. In *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'97)*, München.

Menzel, Wolfgang. 1998. Constraint satisfaction for robust parsing of natural language. *Theoretical and Experimental Artificial Intelligence*, 10(1):77–89.

Menzel, Wolfgang and Ingo Schröder. 1999. Error diagnosis for language learning systems. *ReCALL*, (special edition, May 1999):20 – 30.

Oerder, Martin and Hermann Ney. 1993. Wordgraphs: an efficient interface between continuous-speech recognition and language understanding. In *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'93)*, pages II–119 – II–123, Minneapolis.

Rupp, C.J., Jörg Spilker, Martin Klarner, and Karsten L. Worm. 2000. Combining analyses from various parsers. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag, Berlin, Heidelberg, pages 311–320.

Schröder, Ingo. 2002. *Natural Language Parsing with Graded Constraints*. Dissertation, Universität Hamburg, Fachbereich Informatik.

Schröder, Ingo, Wolfgang Menzel, Kilian Foth, and Michael Schulz. 2000. Modeling dependency grammar with restricted constraints. *Traitement Automatique des Langues (T.A.L.)*, 41(1):97–126.

Schulz, Michael. 2000. *Parsen natürlicher Sprache mit gesteuerter lokaler Suche*. Diplomarbeit, Universität Hamburg, Fachbereich Informatik.

Tessiore, Lorenzo and Walther v. Hahn. 2000. Functional validation of a machine interpretation system: Verbmobile. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag, Berlin, Heidelberg, pages 611–631.

Vogel, Stephan, Franz Josef Och, Christof Tillmann, Sonja Nießen an Hassan Sawaf, and Hermann Ney. 2000. Statisical methods for machine translation. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag, Berlin, Heidelberg, pages 377–393.

Voudouris, Christos. 1997. *Guided Local Search for Combinatorial Optimisation Problems*. Ph.D. thesis, Depertment of Computer Science, University of Essex, Colchester, UK.

Wahlster, Wolfgang. 2000. Mobile speech-to-speech translation of spontaneous dialogs: An overview of the final Verbmobil system. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag, Berlin, Heidelberg, pages 3–21.