Context-based error correction for proper names in a large vocabulary domain

Alexandra Klein
Austrian Research Institute
for Artificial Intelligence
Freyung 6/6, A-1010 Vienna, Austria

Harald Trost
Department of Medical Cybernetics
and Artificial Intelligence
Medical University of Vienna
Freyung 6/2, A-1010 Vienna, Austria

Abstract

Natural Language Understanding (NLU) in the domain of news texts needs to be able to handle proper names, as this word category is highly frequent, and the individual terms are very transitional due to the emergence and disappearance of specific topics in the news. Many speech recognition errors in this domain are due to a limited recognition vocabulary or variations in pronunciation. An error correction component can detect and correct errors by relying on knowledge derived from both the interaction history and factual context. For this purpose, factual knowledge about persons, places and relations between these concepts is constantly being collected with Information Extraction (IE) techniques from available news sources, and this knowledge helps identify misrecognized terms as well as construct appropriate queries. For terms, alternatives which are considered better candidates according to the background knowledge can thus be suggested. In the same domain, this approach may also be used for correcting spelling errors of proper names in written user requests.

1 Introduction

Even with the increasing quality of speech recognition systems, the recognition of proper names still remains a major challenge. This is largely due to two reasons: First, proper names tend to contain unusual spelling and pronunciation. Second, it is difficult to constantly update a spelling and pronunciation dictionary for proper names. The large vocabulary of news texts is particularly difficult to handle for a speech recognition system, mostly due to the high number of proper names which are often only transitional, i.e. they appear and disappear since they are connected to some specific piece of news.

Various difficulties are associated with proper names in speech recognition. In dialogue systems for train table information, for instance, it is necessary to train the system with all the proper names which may appear in the databases. Whenever the system needs to be open for new proper names, e.g. as users need to give their names as input to the system, system designer usually resort to having users spell their names (Seneff et al., 2003), which is not feasible in the domain of news texts. For this domain, we therefore consult background knowledge which is derived from news documents on the web. This background knowledge is constantly updated by means of Information Extraction (IE) mechanisms. Using the derived background knowledge can help detect and correct errors caused by speech recognition due to out-of-vocabulary words, inconsistent pronunciations of foreign names and acoustic similarities of proper names which refer to unrelated objects. This considerably decreases the need for training in the newspaper domain. The derived background knowledge can also be used in the detection and correction of spelling errors in users' written requests for information from news texts.

The system which is described in this paper uses knowledge from the context of the textual bases of the news as well as the context of the users' spoken or written utterances represented as specific goal-oriented user actions (Alexandersson and Reithinger, 1997). This contextual evidence is combined and compared to the results of the speech recognition. If discrepancies are detected, the problems can usually be resolved by relying on the contextual knowledge sources. The same repair mechanisms can be employed for problems due to typological or or-

thographic errors in the written user utterances. In the following, we are first going to describe the overall system and domain as well as the problems with speech recognition which we have encountered. The use of two corpora for utterance understanding are explained: a corpus of spoken and written utterances and a newspaper corpus which has been compiled in the course of more than one year. The following section gives an outline of the context-based error correction as it is employed in the system, based on extracted concepts and relations from the newspaper corpus and on evidence from the utterance history. This knowledge also contributes to query expansion, as is described in the next section. A conclusion will refer to further work, which includes the automatic extraction of the patterns themselves.

2 System and domain

The domain of the system described in this paper are newspaper texts. Users can ask written and spoken questions regarding the content of newspaper archives. They determine the functionality of the input modes themselves, by freely combining search requests with browser commands in spoken or typed utterances. The development of the system has been backed by empirical data; a Wizard-of-Oz study simulated user behavior. A corpus of Austrian newspaper texts (equaling a year's worth of articles) was collected, the corpus now serves both as a search space and a knowledge base for entities and their relations which are typically searched for by users. In particular, this knowledge base can be used for error correction and query expansion. Figure 1 gives an overview of the architecture.

2.1 Contrastive functionality

For scenarios where spoken as well as written input can be used, little is known about mode preference and coordination. Generally, it is agreed that a contrastive functionality (Oviatt and Olsen, 1994) of several input modes including speech results in the highest performance as well as user acceptance since the naturalness and effectiveness of human language is based on its flexibility to support every possible way of interaction. Direct manipulation is still regarded as the most efficient, natural and intuitive way of interacting with objects on the screen, and mouse clicks are thus often used for simple tasks like following a link or moving between two documents. Typed input is often used for search terms as this medium is less error-prone than spoken language (in spite of typological errors). Spoken language tends to be employed for complex queries as they can be expressed quickly in this medium. Of course, this further complicates the task of interpreting users' spoken utterances, which often contain errors and other phenomena of spontaneous speech.

2.2 Typology of user actions

For a better interpretation of the various user requests, we have decided to distinguish between three types of user actions in the evaluation of the multimodal corpus:

- Combined queries referring to form and content, i.e. utterances which contain browsing requests as well as search terms, e.g. *back to the article about Stephan Eberharter in the Sports section.*
- Content queries containing only search terms, e.g. *information about the Olympic Games.*
- Browsing requests and metalanguage, e.g. *back, I want to start a new search, I think I made a mistake I want to start again.*

3 Speech recognition errors in the newspaper domain

Speech recognition errors occur most often in queries which contain search terms; this is due to the high rate of proper names among search terms: Newspaper texts tend to contain many foreign names of people, places and also sometimes titles or functions. This leads to a high word error rate for speech recognition in this domain. The error rate can be reduced by means of additional training, but it tends to remain considerable. In order to assess the error rates, 3,816 noun phrases containing at least one proper name were read to a commercial speech recognition system. 1,723 (45.15%) of them contained at least one error in the recognition result rated best by the system although the vocabulary had been completely trained for the speaker-dependent speech recognition system.

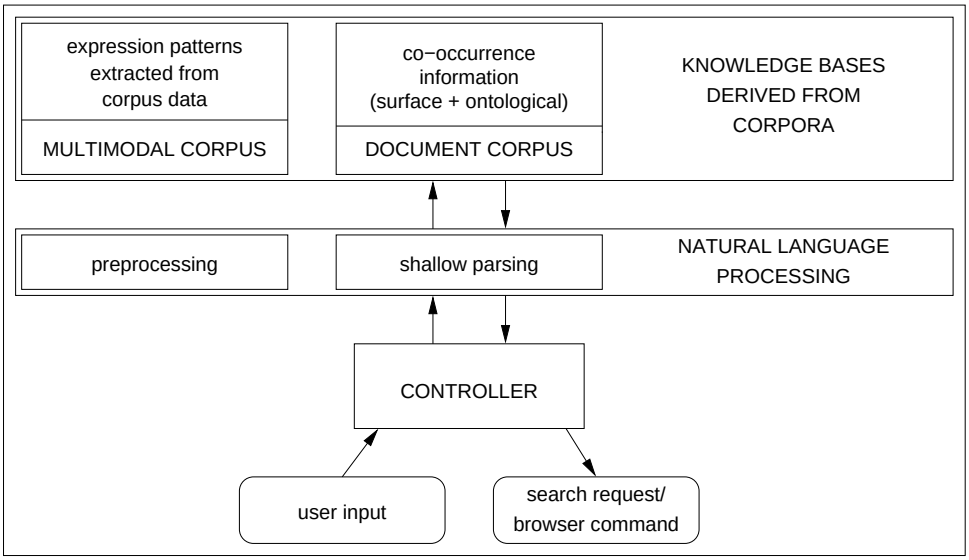

Figure 1: *Architecture and knowledge sources.*

Many errors in the newspaper domain stem from the fact that proper names in the utterances are not contained in the vocabulary of the speech recognizer. Some approaches, e.g. (Gurevych et al., 2003), correct errors by processing n-best lists and generating a new set of alternative, more semantically likely speech recognition hypotheses based on a manually built semantic dictionary. In contrast, the work described in this paper operates on the best hypothesis and detects and corrects errors based on the background knowledge automatically derived from the corpus and the discourse context.

By drawing on this background knowledge, it is possible to identify error candidates even if the correct alternatives are not contained in the speech recognition vocabulary or if they were pronounced in way which had not been anticipated. In these cases, the correct terms are usually not contained in the word graphs, and it is useful to consult other knowledge sources.

4 Spoken corpus

In order to obtain a corpus of typical user actions, we carried out Wizard-of-Oz (WoZ) (Fraser and Gilbert, 1991) experiments. 43 test persons were presented with twelve tasks each, with the tasks being assigned different types and combinations of input modes: spoken input, typed input and free typed or spoken input (Klein et al., 2001). Both in the spoken and written input modes, the test persons were encouraged to mix browser commands with the actual search requests. According to their experience with the web, the users were grouped as experts and non-experts. During the experiments, it turned out that the non-experts in particular liked to combine browser commands with actual search requests, i.e. *back to hockey*. For the experiments, predefined web pages were used in order to achieve comparable results. We simulated perfect speech recognition although we are well aware of the problems associated with WoZ experiments in a speech environment, particularly the simulation of speech recognition errors.

Results show that experts as well as non-experts prefer multimodal interaction over single input modes, and non-experts in particular were able speed up task completion times significantly with a combination of spoken and written input with mouse clicks. The subjects' utterances were recorded, and typed input and mouse clicks were logged and later aligned with the spoken input. The corpus gave us a good impression of how different user actions are lexicalized, how people prefer to refer to dates and newspapers, and how they combine the different elements in their utterances. Furthermore, the data from the experiments helped gain insight into spontaneous-speech phenomena which are likely to be encountered in the scenario. Users' utterances contained numerous hesitations, cor-

rections, meta-comments concerning the system which were mixed with the actual search requests.

5 Newspaper corpus

In order to support the analysis of the search terms in the users' queries, we have compiled a written corpus which contains a year's worth of articles from the Austrian daily newspaper *Der Standard*. This corpus serves as a base for error correction and 'query expansion' in the sense that users' queries can be complemented with information lost during speech recognition, omitted by the users or linked to some contextual information neglected in the analysis, cf. (Klein and Trost, 2003). Apart from a robust natural-language understanding component, a sound representation of context and world knowledge considerably facilitates the interpretation of user requests, as described in (Klein et al., 2002). In our system, the document corpus serves as a foundation for deriving ontological knowledge which resolves contextual ambiguities and contradictions in the user utterances.

At present, the corpus consists of 6,738,284 words, and new texts are added on a daily base. A segment which at the moment comprises 1,398,732 words of this corpus has been tagged automatically. Of these words, 1,137,544 words were tagged as proper names, which makes up 81.33% of the segment: a considerable portion, which is a typical feature of news texts.

6 Context-based error correction

Context-based error correction relies on knowledge about the factual and the discourse contexts of terms and entities mentioned in the user utterances with respect to the background knowledge of the system. Many speech recognition errors concern proper names, and proper names tend to be contained in noun phrases concerning persons, places or other geographical information, as well as companies, political parties etc. Often, prepositions describe the relation between entities referred to by proper names. Specific patterns also refer to the relationship between entities. In news texts, these patterns tend to be ritualized and frequent in describing specific relations. It was therefore possible in our system to e.g. extract 43 capital-country relations with only a few patterns yet without any errors. Instances of relations can be extracted constantly whenever they occur. This way, the most feasible concepts and terms are added to the background knowledge.

6.1 Extraction of concepts and relations

For error correction as well as query expansion, the system relies on Information Extraction (IE) techniques (Riloff, 1999; Neumann, 2001). Usually, it is distinguished between Information Retrieval (IR) methods, which search for documents, and IE methods, which search for passages in documents, for finding information searched by a user. Since it was our task to compare the content of the users' utterances to the content of the document corpus and to correct and expand user queries when necessary, Information Extraction techniques can be used of identifying matches between user queries and passages of text, which is a prerequisite for successful error correction and query expansion.

The extraction of text content represented as case frames, concepts, conceptual relations and semantic roles operates on the interface between syntax and semantics. The segmentation of different elements occurring in the text and the rough grouping of these elements require at least partial syntactic processing. Semantics comes into play for the analysis of the type of relations which exist between elements of the conceptual structure. Generally, it is very difficult to discern the type of semantic relation. Recently, there has been renewed interest in this task, cf. e.g. (Meyer and Dale, 2002; Klein and Trost, 2003).

Typically, in IE systems, partial parsing methods are used for identifying noun phrases, and as far as possible, dependencies between phrases and verbs (i.e. case frames) are found. Named entities are then extracted from the noun phrases, and the named entities are categorized as names of people, organizations or companies, locations etc. Since business (Ciravegna et al., 1999), news and medicine are typical IE domains, named-entity extraction and classification are usually concerned with entities relevant for these domains.

For preprocessing the textual data in the news documents, a Brill tagger (Brill, 1992) trained for German on the Negra corpus (Skut

et al., 1997) was used for tagging. The rule-based part-of-speech tagger works by first assigning each word its most likely tag, and then changing word taggings based on contextual cues. The Cass partial parser (Abney, 1996) which is based on cascaded finite-state automata was used for partial parsing. Complex noun phrases were extracted, and special attention was paid to possible named entities. These entities were identified by grammar rules and keywords which can be easily incorporated into the Cass grammar.

The surface forms which have been extracted from texts need to be mapped onto their underlying concepts. Ontologies as they are employed in IE systems are usually concerned with the role of concepts (expressed by words) in a specific context or domain. For this task, ontologies are used (such as WordNet (Fellbaum, 1998)) or derived automatically.

For named entities, we first concentrated on phrases referring to names, organizations and locations. It was possible to extract lists of names (first name and last name), titles and official functions, cities, and countries. While it has not been possible to build a comprehensive knowledge base even from our large newspaper corpus, in many cases it has been possible to derive links between names and functions, between names/functions and countries and between cities and countries. Consequently, the following relations were used:

- `has-function(Name,Function)`
  Example:
  `has-function(Kofi Annan, UN Secretary-General)`
- `lives-in-or-is-from(Name/Function, City/Country)`
  Example:
  `lives-in-or-is-from(Queen Elizabeth, England)`
- `lies-in(City,Country)`
  Example:
  `lies-in(Gouda,Netherlands)`
- is-capital-of(City,Country)
  Example:
  `is-capital-of(Vienna,Austria)`

Along the same lines, we have also automatically built relations concerning political parties, rivers and other geographical landmarks, but this database is much smaller and therefore less useful. Since functions and titles are often combined with an adjective referring to a country, in many cases it is possible to connect names, titles and organizations to countries, e.g. *German foreign minister Joschka Fischer*.

The extracted relations can be combined, and as a result we have obtained a small, automatically built thesaurus. So far, the thesaurus only contains information about people and their functions as well as places, and the relations between people and places. Yet, this information is very useful in dealing with newspaper corpora, and queries of this kind frequently serve as input to newspaper search engines. For the domain of newspaper texts, this corpus can complement WordNet (Fellbaum, 1998) since it has at the same time a broader base than a manually built thesaurus (by covering a large number of documents) and it is simultaneously better customized for our domain (in our case to Austrian newspaper texts).

6.2 Concept candidates for error correction

In order to detect speech-recognition errors in the analysis of spoken utterances, or typological/orthographic errors in written utterances, we basically evaluate the context as it can be mapped to thesaurus relations for the domain of news texts. As it has been mentioned, in addition to the problems associated with out-of vocabulary words, the commercial speech recognition system which we use tends to make errors in name recognition even for names which have previously been trained. Our thesaurus relations help identify recognition errors and correct them by using the phrasal context. E.g. *Prime Minister Ariel Sharon* was recognized as *Prime Minister Ariel Scharrer*. Using the thesaurus relations, the system determines that in the newspaper corpus, *Ariel* appeared only with *Sharon*, which is a good indication that *Ariel Sharon* is meant. A further clue lies in the fact that in the newspaper corpus, there were many cooccurrences of *Ariel Sharon* and *Prime Minister*, which supports the hypothesis that this is who is meant in the user utterance. A further analysis of the

recognition errors could improve correction. So far, we have only experimented with building 'synsets' of errors; for *Sharon* the synset would be `< Sharon, Scharrer, Schoch, Shah um, verwunden >`. Of course, these forms cannot be exchanged automatically; first, an analysis of the context always has to be carried out. It would also be helpful to find a distance measure to compute distances between the recognized and the actual word forms. So far, our attempts to use existing measures have not proven effective, but we intend to continue experimenting with different distance measures. This also applies to typological/orthographic errors, where distance measures have to be adapted according to keyboard constellations and typical orthographic problems.

7 Query expansion

After the user utterance has been reconstructed or approximated, an appropriate search query has to be generated. For this task, the search terms are collected from the user utterances. Since utterances tend to be elliptical, and a search for the mentioned terms often does not yield satisfying results, query expansion is carried out, again based on background knowledge from factual context and the interaction history. Furthermore, coocurrence information is used in modifying search queries for a more accurate and efficient retrieval.

From our corpus of newspaper texts, adjective-noun- and adjective-proper-name pairs were extracted and counted. These pairs were stored and consulted in query interpretation. Since the texts are tagged automatically, the lists of adjectives and nouns/proper names contain a considerable number of errors. Therefore it is helpful to use large amounts of text; it may even be useful to eventually introduce a threshold so that only adjective-noun/proper-name pairs which appear more than once or a certain number of times are considered. This of course can not prevent systematic tagging errors. A robust stemming algorithm maps all adjective-noun/proper-name pairs to an approximate 'stem', thus eliminating inflected forms which result in morphological variation which is typical for the German language.

Whenever a word is encountered in processing which can be considered an adjective, it is kept; whenever the following word may be a noun or a proper name, it is checked whether the adjective-noun/proper-name combination is contained in the repository of adjective-noun/proper-name combinations which has previously been extracted from a corpus. If the adjective-noun/proper-name combination is found, it is passed on to the search engine as a query. Whenever the combination has not occurred in the corpus, only the noun or proper name is considered a key word. Our approach distinguishes noun phrases which have a record of cooccurrence from noun phrases which may be spontaneous expressions or modifications or even errors created by users. For example, the phrase *European countries* would be retained while *participating countries* would be reduced to the noun. Some adjectives used in search expressions serve to qualify the global search expression rather than the noun or proper name in question. For example, a search for *yesterday's speech* would only yield articles from the day after a speech, not about the speech in general.

Our approach to query expansion is in some aspects similar to the work described in (Jourlin et al., 1999). Here, automatic transcriptions of spoken news texts, which contain many errors made in speech recognition, are accessed via written commands, whereas in our case, multimodal including spoken commands refer to written newspaper documents. For query expansion, (Jourlin et al., 1999) used geographic partly ordered sets (posets) with `contain` relations. The relation knowledge base was manually built.

In our approach, names and titles of persons were extracted from the newspaper corpus with the help of a rule-based mechanism, as it has been described in the previous section. If names and titles have a unique relation, i.e. there was only one cooccurrence in the corpus, they are treated as WordNet-like 'synsets'. For query expansion, all elements of the synset are passed on to the search engine. If the relation is not unique, e.g. when a person is linked to a title which is connected also to other names, like 'member of Parliament', query expansion is less helpful, since many documents might turn up in search which refer to other people. Consequently, these instances are more useful for error correction.

8 Conclusion

We have developed an error correction mechanisms for a system which allows users to access news texts on the web. In order to bridge the gap between the potentially unlimited set of proper names in news, which may appear in user queries, and the limited vocabulary of speech recognition systems, and in order to compensate for recognition errors due to foreign words, background knowledge is needed. It allows an embedding of the user query into the discourse history and the factual context. While it is easily conceivable that our error correction mechanism might be improved by using as input word graphs or n-best lists as they are produced by speech recognition systems, at the moment, our approach operates on the best hypothesis. Since a large amount of errors are due to out-of-vocabulary words, this approach already helps correct many errors which occur in the news domain, and it can easily be adapted to different commercial speech recognition systems. It can also be used for correcting typological or orthographic errors which are frequently encountered in written access to news texts.

Consequently, simple patterns provide an efficient means of extracting and representing knowledge about concepts which are crucial in newspaper texts, yet which often cause problems in the interpretation of queries in the news domain due to speech recognition or spelling errors. By means of these patterns, knowledge can constantly be updated as certain proper names appear and disappear in the news text. At the moment, the extraction patterns are static and coded manually; it would be useful to derive them automatically. Furthermore, it would be interesting to complement the findings of the pattern-based approach with the results of systems which rely on distributional phenomena for identifying proper names (e.g. (Borthwick et al., 1998; Cucerzan and Yarowsky, 1999; Valsamidis and Cooper, 1999)).

In our system, the interpretation of multimodal user actions – in our case spoken and written utterances combined with mouse clicks – can be improved by the use of background knowledge derived from corpora and the action history. This allows for greater system usability and robustness. Therefore, the detection and correction of errors during the interpretation of the user requests contributes to a more natural interaction, which – as our empirical study has shown – is a crucial factor in system design.

9 Acknowledgement

The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Education, Science and Culture and the Austrian Federal Ministry for Transport, Innovation and Technology. The project was supported by the Austrian Science Fund (P13704-INF). The authors would like to thank the anonymous reviewers for helpful comments.

References

Steven Abney. 1996. Partial parsing via finite-state cascades. *Natural Language Engineering*, 2(4):337–344.

Jan Alexandersson and Norbert Reithinger. 1997. Learning dialogue structures from a corpus. In *Proceedings of Eurospeech '97*, pages 2231–2235, Rhodes, Greece.

Andrew Borthwick, John Sterling, Eugene Agichtein, and Ralph Grishman. 1998. NYU: Description of the MENE named entity system as used in MUC-7. In *Proceedings of the Seventh Message Understanding Conference (MUC-7)*.

Eric Brill. 1992. A simple rule-based part of speech tagger. In *Proceedings of the Third Conference on Applied Natural Language Processing (ACL)*, Trento, Italy.

F. Ciravegna, A. Lavelli, L. Gilardoni, J. Matiasek, N. Mana, S. Mazza, M. Ferraro, W.J. Black, F. Rinaldi, and D. Mowatt. 1999. FACILE: Classifying texts integrating pattern matching and information extraction. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Stockholm, Sweden*, pages 890–896, Los Altos/Palo Alto/San Francisco. Morgan Kaufmann.

Silviu Cucerzan and David Yarowsky. 1999. Language independent named entity recognition combining morphological and contextual evidence. In *Joint SIGDAT Conference on EMNLP and VLC*.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Norman M. Fraser and G. Nigel Gilbert. 1991. Simulating speech systems. *Computer Speech and Language*, 5(1):81–99.

Iryna Gurevych, Rainer Malaka, Robert Porzel, and Hans-Peter Zorn. 2003. Semantic coherence scoring using an ontology. In *Proceedings of the Joint Human Language Technology and Northern Chapter of the Association for Computational Linguistics Conference (HLT-NAACL)*, pages 88–95, Edmonton, Canada.

Sue E. Jourlin, Pierre Johnson, Karen Spärck Jones, and Philip C. Woodland. 1999. General query expansion techniques for spoken document retrieval. In *Proceedings of the ESCA Workshop on Extracting Information from Spoken Audio*, pages 8–13, Cambridge, UK.

Alexandra Klein and Harald Trost. 2003. Using corpus-based methods for spoken access to news texts on the web. In *Proceedings of Eurospeech'03*, Geneva, Switzerland.

Alexandra Klein, Ingrid Schwank, Michel Généreux, and Harald Trost. 2001. Evaluating Multimodal Input Modes in a Wizard-of-Oz Study for the Domain of Web Search. In Ann Blandford, Jean Vanderdonckt, and Phil Gray, editors, *People and Computer XV – Interaction without Frontiers: Joint Proceedings of HCI 2001 and IHM 2001*, pages 475–483. Springer: London.

Alexandra Klein, Puig-Waldmüller, and Harald Trost. 2002. Robust interpretation of user requests for text retrieval in a multimodal environment. In *COLING 2002: Proceedings of the 19th International Conference on Computational Linguistics*, pages 1233–1237, Taipei, Taiwan.

Josef Meyer and Robert Dale. 2002. Mining a corpus to support associative anaphora resolution. In *Proceedings of the 4th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC 2002)*, Lisbon, Portugal.

Günter Neumann. 2001. Informationsextraktion. In R. Klabunde, editor, *Computerlinguistik und Sprachtechnologie: Eine Einführung*. Spektrum Akademischer Verlag, Heidelberg.

Sharon L. Oviatt and Erik Olsen. 1994. Integration Themes in Multimodal Human-Computer Interaction. In *Proceedings of the ICSLP*, volume 2, pages 551–554, Yokohama. Acoustical Society of Japan.

Ellen Riloff, 1999. *Understanding Language Understanding: Computational Models of Reading, edited by Ashwin Ram and Kenneth Moorman*, chapter Information Extraction as a Stepping Stone toward Story Understanding. MIT Press.

Stephanie Seneff, Grace Chung, and Chao Wang. 2003. Empowering end users to personalize dialogue systems through spoken interaction. In *Proceedings of the 8th European Conference on Speech Communication and Technology*, Geneva, Switzerland.

Wojciech Skut, Brigitte Krenn, Thorsten Brants, and Hans Uszkoreit. 1997. An annotation scheme for free word order languages. In *Proceedings of the Fifth Conference on Applied Natural Language Processing (ANLP-97)*, Washington, DC.

T. Valsamidis and M. Cooper. 1999. An analysis of proper name distributions in italian and french news stories. Moms Propres, ATALA, Paris, May.