Context Management with Topics for Spoken Dialogue Systems

Kristiina Jokinen and **Hideki Tanaka** and **Akio Yokoo**
ATR Interpreting Telecommunications Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-02 Japan
email: {kjokinen|tanakah|ayokoo}@itl.atr.co.jp

Abstract

In this paper we discuss the use of discourse context in spoken dialogue systems and argue that the knowledge of the domain, modelled with the help of dialogue topics is important in maintaining robustness of the system and improving recognition accuracy of spoken utterances. We propose a topic model which consists of a domain model, structured into a topic tree, and the Predict-Support algorithm which assigns topics to utterances on the basis of the topic transitions described in the topic tree and the words recognized in the input utterance. The algorithm uses a probabilistic topic type tree and mutual information between the words and different topic types, and gives recognition accuracy of 78.68% and precision of 74.64%. This makes our topic model highly comparable to discourse models which are based on recognizing dialogue acts.

1 Introduction

One of the fragile points in integrated spoken language systems is the erroneous analyses of the initial speech input.[1] The output of a speech recognizer has direct influence on the performance of other modules of the system (dealing with dialogue management, translation, database search, response planning, etc.), and the initial inaccuracy usually gets accumulated in the later stages of processing. Performance of speech recognizers can be improved by tuning their language model and lexicon, but problems still remain with the erroneous ranking of the best paths: information content of the selected utterances may be wrong. It is thus essential to use contextual information to compensate various errors in the output, to provide expectations of what will be said next and to help to determine the appropriate dialogue state.

However, negative effects of an inaccurate context have also been noted: cumulative error in discourse context drags performance of the system below the rates it would achieve were contextual information

not used (Qu et al., 1996; Church and Gale, 1991). Successful use of context thus presupposes appropriate context management: (1) features that define the context are relevant for the processing task, and (2) construction of the context is accurate.

In this paper we argue in favour of using one type of contextual information, topic information, to maintain robustness of a spoken language system. Our model deals with the *information content* of utterances, and defines the context in terms of *topic types*, related to the current domain knowledge and represented in the form of a topic tree. To update the context with topics we introduce the Predict-Support algorithm which selects utterance topics on the basis of topic transitions described in the topic tree and words recognized in the current utterance. At present, the algorithm is designed as a filter which re-orders the candidates produced by the speech recognizer, but future work encompasses integration of the algorithm into a language model and actual speech recognition process.

The paper is organised as follows. Section 2 reviews the related previous research and sets out our starting point. Section 3 presents the topic model and the Predict-Support algorithm, and section 4 gives results of the experiments conducted with the model. Finally, section 5 summarises the properties of the topic model, and points to future research.

2 Previous research

Previous research on using contextual information in spoken language systems has mainly dealt with speech acts (Nagata and Morimoto, 1994; Reithinger and Maier, 1995; Möller, 1996). In dialogue systems, speech acts seem to provide a reasonable first approximation of the utterance meaning: they abstract over possible linguistic realisations and, dealing with the illocutionary force of utterances, can also be regarded as a domain-independent aspect of communication.[2]

---

[1] Alexandersson (1996) remarks that with a 3000 word lexicon, a 75 % word accuracy means that in practice the word lattice does not contain the actually spoken sentence.

[2] Of course, most dialogue systems include domain dependent acts to cope with the particular requirements of the domain, cf. Alexandersson (1996). Speech acts are also related to the task: information providing, appointment negotiat-

However, speech acts concern a rather abstract level of utterance modelling: they represent the speakers' intentions, but ignore the semantic content of the utterance. Consequently, context models which use only speech act information tend to be less specific and hence less accurate. Nagata and Morimoto (1994) report prediction accuracy of 61.7 %, 77.5 % and 85.1 % for the first, second and third best dialogue act (in their terminology: Illocutionary Force Type) prediction, respectively, while Reithinger and Maier (1995) report the corresponding accuracy rates as 40.28 %, 59.62 % and 71.93 %, respectively. The latter used structurally varied dialogues in their tests and noted that deviations from the defined dialogue structures made the recognition accuracy drop drastically.

To overcome prediction inaccuracies, speech act based context models are accompanied with the information about the task or the actual words used. Reithinger and Maier (1995) describe plan-based repairs, while Möller (1996) argues in favour of domain knowledge. Qu et al. (1996) show that to minimize cumulative contextual errors, the best method, with 71.3% accuracy, is the Jumping Context approach which relies on syntactic and semantic information of the input utterance rather than strict prediction of dialogue act sequences. Recently also keyword-based topic identification has been applied to dialogue move (dialogue act) recognition (Garner, 1997).

Our goal is to build a context model for a spoken dialogue system, and we emphasise especially the system's *robustness*, i.e. its capability to produce reliable and meaningful responses in presence of various errors, disfluencies, unexpected input and out-of-domain utterances, etc. (which are especially notorious when dealing with spontaneous speech). The model is used to improve word recognition accuracy, and it should also provide a useful basis for other system modules.

However, we do not aim at robustness on a merely mechanical level of matching correct words, but rather, on the level of maintaining the *information content* of the utterances. Despite the vagueness of such a term, we believe that speech act based context models are less robust due to the fact that the information content of the utterances is ignored. Consistency of the information exchanged in (task-oriented) conversations is one of the main sources for dialogue coherence, and so pertinent in the context management besides speech acts. Deviations from a predefined dialogue structure, multifunctionality of utterances, various side-sequences, disfluencies, etc. cannot be dealt with on a purely abstract level of illocution, but require knowledge of the domain, expressed in the semantic content of the utterances.

Moreover, in multilingual applications, like speech-to-speech translation systems, the semantic content of utterances plays an important role and an integrated system must also produce a semantic analysis of the input utterance. Although the goal may be a shallow understanding only, it is not enough that the system knows that the speaker uttered a "request": the type of the request is also crucial.

We thus reckon that appropriate context management should provide descriptions of what is said, and that the recognition of the utterance topic is an important task of spoken dialogue systems.

3 The Topic Model

In AI-based dialogue modelling, topics are associated with a particular discourse entity, *focus*, which is currently in the centre of attention and which the participants want to focus their actions on, e.g. Grosz and Sidner (1986). The topic (focus) is a means to describe thematically coherent discourse structure, and its use has been mainly supported by arguments regarding anaphora resolution and processing effort (search space limits). Our goal is to use topic information in predicting likely content of the next utterance, and thus we are more interested in the topic *types* that describe the information conveyed by utterances than the actual topic entity. Consequently, instead of tracing salient entities in the dialogue and providing heuristics for different shifts of attention, we seek a formalisation of the information structure of utterances in terms of the *new information* that is exchanged in the course of the dialogue.

The purpose of our topic model is to assist speech processing, and so extensive and elaborated reasoning about plans and world knowledge is not available. Instead a model that relies on observed facts (= word tokens) and uses statistical information is preferred. We also expect the topic model to be general and extendable, so that if it is to be applied to a different domain, or more factors in the recognition of the information structure of the utterances[3] are to be taken into account, the model could easily adapt to these changes.

The topic model consists of the following parts:

1. domain knowledge structured into a topic tree

2. prior probabilities of different topic shifts

3. topic vectors describing the mutual information between words and topic types

4. Predict-Support algorithm to measure similarity between the predicted topics and the topics supported by the input utterance.

Below we describe each item in detail.

---

ion, argumentation etc. have different communicative purposes which are reflected in the set of necessary speech acts.

[3]For instance, sentential stress and pitch accent are important in recognizing topics in spontaneous speech.
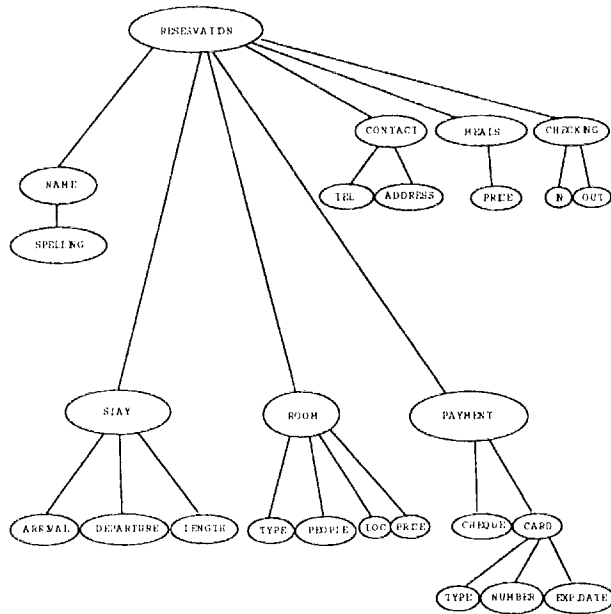
Figure 1: A partial topic tree.

3.1 Topic trees

Originally "focus trees" were proposed by (McCoy and Cheng, 1991) to trace foci in NL generation systems. The branches of the tree describe what sort of shifts are cognitively easy to process and can be expected to occur in dialogues: random jumps from one branch to another are not very likely to occur, and if they do, they should be appropriately marked. The focus tree is a subgraph of the world knowledge, built in the course of the discourse on the basis of the utterances that have occurred. The tree both constrains and enables prediction of what is likely to be talked about next, and provides a top-down approach to dialogue coherence.

Our topic tree is an organisation of the domain knowledge in terms of topic types, bearing resemblance to the topic tree of Carcagno and Iordanskaja (1993). The nodes of the tree[4] correspond to topic types which represent clusters of the words expected to occur at a particular point of the dialogue. Figure 1 shows a partial topic tree in a hotel reservation domain.

For our experiments, topic trees were hand-coded from our dialogue corpus. Since this is time-consuming and subjective, an automatic clustering program, using the notion of a topic-binder, is currently under development.

Our corpus contains 80 dialogues from the bilingual ATR Spoken Language Dialogue Database.

---

[4] We will continue talking about a topic *tree*, although in statistical modelling, the tree becomes a topic *network* where the shift probability between nodes which are not daughters or sisters of each other is close to zero.

The dialogues deal with hotel reservation and tourist information, and the total number of utterances is 4228. (Segmentation is based on the information structure so that one utterance contains only one piece of new information.) The number of different word tokens is 27058, giving an average utterance length 6,4 words.

The corpus is tagged with speech acts, using a surface pattern oriented speech act classification of Seligman et al. (1994), and with topic types. The topics are assigned to utterances on the basis of the new information carried by the utterance. New information (Clark and Haviland, 1977; Vallduví and Engdahl, 1996) is the locus of information related to the sentential nuclear stress, and identified in regard to the previous context as the piece of information with which the context is updated after uttering the utterance. Often new information includes the verb and the following noun phrase.

More than one third of the utterances (1747) contain short fixed phrases (*Let me confirm; thank you; good-bye; ok; yes*), and temporizers (*well, ah, uhm*). These utterances do not request or provide information about the domain, but control the dialogue in terms of time management requests or conventionalised dialogue acts (feedback-acknowledgements, thanks, greetings, closings, etc.) The special topic type IAM, is assigned to these utterances to signify their role in InterAction Management. The topic type MIX is reserved for utterances which contain information not directly related to the domain (safety of the downtown area, business taking longer than expected, a friend coming for a visit etc.), thus marking out-of-domain utterances. Typically these utterances give the reason for the request.

The number of topic types in the corpus is 62. Given the small size of the corpus, this was considered too big to be used successfully in statistical calculations, and they were pruned on the basis of the topic tree: only the topmost nodes were taken into account and the subtopics merged into approproate mother topics. Figure 2 lists the pruned topic types and their frequencies in the corpus.

tag	count	%	interpretation
iam	1747	41.3	Interaction Management
room	826	19.5	Room, its properties
stay	332	7.9	Staying period
name	320	7.6	Name, spelling
res	310	7.3	Make/change/extend/ cancel reservation
paym	250	5.9	Payment method
contact	237	5.6	Contact Info
meals	135	3.2	Meals (breakfast, dinner)
mix	71	1.7	Single unique topics

Figure 2: Topic tags for the experiment.

3.2 Topic shifts

On the basis of the tagged dialogue corpus, probabilities of different topic shifts were estimated. We used the Carnegie Mellon Statistical Language Modeling (CMU SLM) Toolkit, (Clarkson and Rosenfeld, 1997) to calculate probabilities. This builds a trigram backoff model where the conditional probablilities are calculated as follows:

$$p(w3|w1,w2) =$$

$$\begin{cases} p3(w1,w2,w3) & \text{if trigram exists} \\ bo\_wt2(w1,w2) \times p(w3|w2) & \text{if bigram (w1,w2)} \\ & \text{exists} \\ p(w3|w2) & \text{otherwise.} \end{cases}$$

$$p(w2|w1) =$$

$$\begin{cases} p2(w1,w2) & \text{if bigram exists} \\ bo\_wt1(w1) \times p1(w2) & \text{otherwise.} \end{cases}$$

3.3 Topic vectors

Each word type may support several topics. For instance, the occurrence of the word *room* in the utterance *I'd like to make a room reservation.* supports the topic MAKERESERVATION, but in the utterance *We have only twin rooms available on the 15th.* it supports the topic ROOM. To estimate how well the words support the different topic types, we measured *mutual information* between each word and the topic types. Mutual information describes how much information a word w gives about a topic type t, and is calculated as follows (ln is log base two, $p(t|w)$ the conditional probability of t given w, and $p(t)$ the probability of t):

$$I(w,t) = \ln \frac{p(w,t)}{p(w) \cdot p(t)} = \ln \frac{p(t|w)}{p(t)}$$

If a word and a topic are negatively correlated, mutual information is negative: the word signals absence of the topic rather than supports its presence. Compared with a simple counting whether the word occurs with a topic or not, mutual information thus gives a sophisticated and intuitively appealing method for describing the interdependence between words and the different topic types.

Each word is associated with a *topic vector*, which describes how much information the word w carries about each possible topic type t_i:

$$topvector(mi(w,t_1), mi(w,t_2), ..., mi(w,t_n))$$

For instance, the topic vector of the word *room* is:

```
topvector(room,[mi(0.21409750769169117,contact),
          mi(-5.5258041314543815,iam),
          mi(-3.831955835588453,meals),
          mi(0,mix),
          mi(-1.26971341136738,name),
          mi(-2.720924523199709,paym),
          mi(0.9687353561881407,res),
          mi(1.9035899442740105,room),
          mi(-4.130179669884547,stay)]).
```

The word supports the topics ROOM and MAKE-RESERVATION (res), but gives no information about MIX (out-of-domain) topics, and its presence is highly indicative that the utterance is not at least IAM or STAY. It also supports CONTACT because the corpus contains utterances like *I'm in room 213* which give information about how to contact the customer who is staying at a hotel.

The topic vectors are formed from the corpus. We assume that the words are independently related to the topic types, although in the case of natural language utterances this may be too strong a constraint.

3.4 The Predict-Support Algorithm

Topics are assigned to utterances given the previous topic sequence (what has been talked about) and the words that carry new information (what is actually said). The Predict-Support Algorithm goes as follows:

1. Prediction: get the set of likely next topics in regard to the previous topic sequences using the topic shift model.

2. Support: link each NewInfo word w_j of the input to the possible topics types by retrieving its topic vector. For each topic type t_i, add up the amounts of mutual information $mi(w_j;t_i)$ by which it is supported by the words w_j, and rank the topic types in the descending order of mutual information.

3. Selection:

   (a) Default: From the set of predicted topics, select the most supported topic as the current topic.

   (b) What-is-said heuristics: If the predicted topics do not include the supported topic, rely on what is said, and select the most supported topic as the current topic (cf. the Jumping Context approach in Qu et al. (1996)).

   (c) What-is-talked-about heuristics: If the words do not support any topic (e.g. all the words are unknown or out-of-domain), rely on what is predicted and select the most likely topic as the current topic.

Figure 3 shows schematically how the algorithm works.

$$U_1 = w_{11}, w_{12}, ..., w_{1m} \longrightarrow T_1$$
$$U_2 = w_{21}, w_{22}, ..., w_{2m} \longrightarrow T_2$$
$$U_3 = w_{31}, w_{32}, ..., w_{3m} \longrightarrow T_3$$

Prediction:
$$T_n = \max_{T_k} p(T_k | T_{k-2} T_{k-1})$$

$$U_n = w_{n1}, w_{n2}, ..., w_{nm} \longrightarrow T_n$$

$mi(W_{n1}, T_a)$	$mi(W_{n2}, T_a)$	$\cdots$	$mi(W_{nm}, T_a)$
$mi(W_{n1}, T_b)$	$mi(W_{n2}, T_b)$	$\cdots$	$mi(W_{nm}, T_b)$
$mi(W_{n1}, T_k)$	$mi(W_{n2}, T_k)$	$\cdots$	$mi(W_{nm}, T_k)$

Support:

$$mi(U_n, T_k) = \sum_{i=1}^{m} mi(W_{ni}, T_k) \qquad T_n = \max_{T_k} mi(U_n, T_k)$$

Select:

Default: $\quad T_n = \max_{T_k} mi(U_n, T_k) \quad$ and $\quad T_n = \max_{T_k} p(T_k | T_{k-2} T_{k-1})$

What is said: $\quad T_n = \max_{T_k} mi(U_n, T_k)$

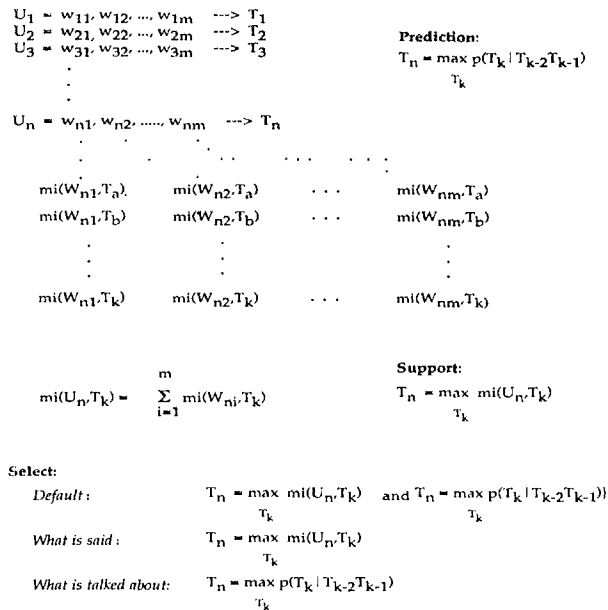*What is talked about:* $\quad T_n = \max_{T_k} p(T_k | T_{k-2} T_{k-1})$

Figure 3: Scheme of the Predict-Support Algorithm.

Using the probabilities obtained by the trigram backoff model, the set of likely topics is actually a set of all topic types ordered according to their likelihood. However, the original idea of the topic trees is to constrain topic shifts (transitions from a node to its daughters or sisters are favoured, while shifts to nodes in separate branches are less likely to occur unless the information under the current node is exhaustively discussed), and to maintain this restrictive property, we take into consideration only topics which have probability greater than an arbitrary limit p.

Instead of having only one utterance analysed at the time and predicting its topic, a speech recognizer produces a word lattice, and the topic is to be selected among candidates for several word strings. We envisage the Predict-Support algorithm will work in the described way in these cases as well. However, an extra step must be added in the selection process: once the topics are decided for the n-best word strings in the lattice, the current topic is selected among the topic candidates as the highest supported topic. Consequently, the word string associated with the selected topic is then picked up as the current utterance.

We must make two caveats for the performance of the algorithm, related to the sparse data problem in calculating mutual information. First, there is no difference between out-of-domain words and unknown but in-domain words: both are treated as providing no information about the topic types. If such words are rare, the algorithm works fine since the other words in the utterance usually support the correct topic. However, if such words occur fre-

quently, there is a difference in regard to whether the unknown words belong to the domain or not. Repeated out-of-domain words may signal a shift to a new topic: the speaker has simply jumped into a different domain. Since the out-of-domain words do not contribute to any expected topic type, the topic shift is not detected. On the other hand, if unknown but in-domain words are repeated, mutual information by which the topic types are supported is too coarse and fails to make necessary distinctions; hence, incorrect topics can be assigned. For instance, if *lunch* is an unknown word, the utterance *Is lunch included?* may get an incorrect topic type ROOMPRICE since this is supported by the other words of the utterance whose topic vectors were build on the basis of the training corpus examples like *Is tax included?*

The other caveat is opposite to unknown words. If a word occurs in the corpus but only with a particular topic type, mutual information between the word and the topic becomes high, while it is zero with the other topics. This co-occurrence may just be an accidental fact due to a small training corpus, and the word can indeed occur with other topic types too. In these cases it is possible that the algorithm may go wrong: if none of the predicted topics of the utterance is supported by the words, we rely on the What-is-said heuristics and assign the highly supported but incorrect topic to the utterance. For instance, if *included* has occurred only with ROOM-PRICE, the utterance *Is lunch included?* may still get an incorrect topic, even though *lunch* is a known word: mutual information $mi(included, RoomPrice)$ may be greater than $mi(lunch, Meals)$.

4 Experiments

We tested the Predict-Support algorithm using cross-validation on our corpus. The accuracy results of the first predictions are given in Table 4. PP is the corpus perplexity which represents the average branching factor of the corpus, or the number of alternatives from which to choose the correct label at a given point.

For the pruned topic types, we reserved 10 randomly picked dialogues for testing (each test file contained about 400-500 test utterances), and used the other 70 dialogues for training in each test cycle. The average accuracy rate, 78.68 % is a satisfactory result. We also did another set of cross-validation tests using 75 dialogues for training and 5 dialogues for testing, and as expected, a bigger training corpus gives better recognition results when perplexity stays the same.

To compare how much difference a bigger number of topic tags makes to the results, we conducted cross-validation tests with the original 62 topic types. A finer set of topic tags does worsen

Test type	PP	PS-algorithm	BO model
Topics = 10 train = 70 files	3.82	78.68	41.30
Topics = 10 train = 75 files	3.74	80.55	40.33
Topics = 62 train = 70 files	5.59	64.96	41.32
Dacts = 32 train = 70 files	6.22	58.52	19.80

Figure 4: Accuracy results of the first predictions.

the accuracy, but not as much as we expected: the Support-part of the algorithm effectively remedies prediction inaccuracies.

Since the same corpus is also tagged with speech acts, we conducted similar cross-validation tests with speech act labels. The recognition rates are worse than those of the 62 topic types, although perplexity is almost the same. We believe that this is because speech acts ignore the actual content of the utterance. Although our speech act labels are surface-oriented, they correlate with only a few fixed phrases (*I would like to; please*), and are thus less suitable to convey the semantic focus of the utterances, expressed by the content words than topics, which by definition deal with the content.

As the lower-bound experiments we conducted cross-validation tests using the trigram backoff-model, i.e. relying only on the context which records the history of topic types. For the first ranked predictions the accuracy rate is about 40%, which is on the same level as the first ranked speech act predictions reported in Reithinger and Maier (1995).

The average precision of the Predict-Support algorithm is also calculated (Table 5). Precision is the ratio of correctly assigned tags to the total number of assigned tags. The average precision for all the pruned topic types is 74.64%, varying from 95.63% for ROOM to 37.63% for MIX. If MIX is left out, the average precision is 79.27%. The poor precision for MIX is due to the unknown word problem with mutual information.

Topic type	Precision	Topic type	Precision
contact	55.75	paym	83.25
iam	79.13	res	62.13
meals	82.13	room	95.63
name	88.12	stay	88.00
mix	37.63	Average	74.64

Figure 5: Precision results for different topic types.

The results of the topic recognition show that the model performs well, and we notice a considerable improvement in the accuracy rates compared to accuracy rates in speech act recognition cited in section 2 (modulo perplexity). Although the rates are somewhat optimistic as we used transcribed dialogues (=

the correct recognizer output), we can still safely conclude that topic information provides a promising starting point in attempts to provide an accurate context for the spoken dialogue systems. This can be further verified in the perplexity measures for the *word* recognition: compared to a general language model trained on non-tagged dialogues, perplexity decreases by 20 % for a language model which is trained on topic-dependent dialogues, and by 14 % if we use an open test with unknown words included as well (Jokinen and Morimoto, 1997).

At the end we have to make a remark concerning the relevance of speech acts: our argumentation is not meant to underestimate their use for other purposes in dialogue modelling, but rather, to emphasise the role of topic information in successful context management: in our opinion the topics provide a more reliable and straighforward approximation of the utterance meaning than speech acts, and should not be ignored in the definition of context models for spoken dialogue systems.

5 Conclusions

The paper has presented a probabilistic topic model to be used as a context model for spoken dialogue systems. The model combines both top-down and bottom-up approaches to topic modelling: the topic tree, which structures domain knowledge, provides expectations of likely topic shifts, whereas the information structure of the utterances is linked to the topic types via topic vectors which describe mutual information between the words and topic types. The Predict-Support Algorithm assigns topics to utterances, and achieves an accuracy rate of 78.68 %, and a precision rate of 74.64%.

The paper also suggests that the context needed to maintain robustness of spoken dialogue systems can be defined in terms of *topic types* rather than speech acts. Our model uses actually occurring words and topic information of the domain, and gives highly competitive results for the first ranked topic prediction: there is no need to resort to extra information to disambiguate the three best candidates. Construction of the context, necessary to improve word recognition and for further processing, becomes thus more accurate and reliable.

Research on statistical topic modelling and combining topic information with spoken language systems is still new and contains several aspects for future research. We have mentioned automatic domain modelling, in which clustering methods can be used to build necessary topic trees. Another research issue is the coverage of topic trees. Topic trees can be generalised in regard to world knowledge, but this requires deep analysis of the utterance meaning, and an inference mechanism to reason on conceptual relations. We will explore possibilities to

extract semantic categories from the parse tree and integrate these with the topic knowledge. We will also investigate further the relation between topics and speech acts, and specify their respective roles in context management for spoken dialogue systems. Finally, statistical modelling is prone to sparse data problems, and we need to consider ways to overcome inaccuracies in calculating mutual information.

References

J. Alexandersson. 1996. Some ideas for the automatic acquisition of dialogue structure. In *Dialogue Management in Natural Language Processing Systems*, pages 149–158. Proceedings of the 11th Twente Workshop on Language Technology, Twente.

D. Carcagno and Lidija Iordanskaja. 1993. Content determination and text structuring: two interrelated processes. In H. Horacek and M. Zock, editors, *New Concepts in Natural Language Generation*, pages 10–26. Pinter Publishers, London.

K. W. Church and W. A. Gale. 1991. Probability scoring for spelling correction. *Statistics and Computing*, (1):93–103.

H. H. Clark and S. E. Haviland. 1977. Comprehension and the given-new contract. In R. O. Freedle, editor, *Discourse Production and Comprehension, Vol.1*. Ablex.

P. Clarkson and R. Rosenfeld. 1997. Statistical language modeling using the CMU-Cambridge toolkit. In *Eurospeech-97*, pages 2707–2710.

P. Garner. 1997. On topic identification and dialogue move recognition. *Computer Speech and Language*, 11:275–306.

B. J. Grosz and C. L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

K. Jokinen and T. Morimoto. 1997. Topic information and spoken dialogue systems. In *NLPRS-97*, pages 429–434. Proceedings of the Natural Language Processing Pacific Rim Symposium 1997, Phuket, Thailand.

K. McCoy and J. Cheng. 1991. Focus of attention: Constraining what can be said next. In C. L. Paris, W. R. Swartout, and W. C. Moore, editors, *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, pages 103–124. Kluwer Academic Publishers, Norwell, Massachusetts.

J-U. Möller. 1996. Using DIA-MOLE for unsupervised learning of domain specific dialogue acts from spontaneous language. Technical Report FBI-HH-B-191/96, University of Hamburg.

M. Nagata and T. Morimoto. 1994. An information-theoretic model of discourse for next utterance type prediction. In *Transactions of Information Processing Society of Japan*, volume 35:6, pages 1050–1061.

Y. Qu, B. Di Eugenio, A. Lavie, L. Levin, and C. P. Rosè. 1996. Minimizing cumulative error in discourse context. In *Dialogue Processing in Spoken Dialogue Systems*, pages 60–64. Proceedings of the ECAI'96 Workshop, Budapest, Hungary.

N. Reithinger and E. Maier. 1995. Utilizing statistical dialogue act processing in verbmobil. In *Proceedings of the 33rd Annual Meeting of the ACL*, pages 116–121.

M. Seligman, L. Fais, and M. Tomokiyo. 1994. A bilingual set of communicative act labels for spontaneous dialogues. Technical Report ATR Technical Report TR-IT-81, ATR Interpreting Telecommunications Research Laboratories, Kyoto, Japan.

E. Vallduví and E. Engdahl. 1996. The linguistic realization of information packaging. *Linguistics*, 34:459–519.