Automatic Induction of Dialogue Structure from the Companions Dialogue Corpus

Name1 Surname1[1] and **Name2 Surname2**[2] and **Name3 Surname3**[3] and **Name4 Surname4**[4]

1 Introduction

Text Tiling [7] [9] and 'lexical chain' [10],[2] techniques have been used in the field of text summarisation to identify segments of prose that are *about* the same topic. We have used related techniques but applied them to a dialogue corpus rather than to prose, simultaneously combining them with an automatically derived analysis of the functional structure of the dialogue (in terms of 'dialogue acts'), and for a quite different purpose. By analysing what each dialogue turn is *about*, combining this information with a concise characterisation of the speaker's intention (a coarse-grained 'dialogue act'), and analysing the sequential patterns of these pairs, we have discovered significant segments of dialogue structure—multi-act dialogue moves, which are applicable across domains, tasks and conversational styles, and which we call 'dialogue action frameworks'. [DGF: DO WE WANT TO SAY WE HAVE DONE THIS WORK AND FOUND THIS THING, OR THAT WE ARE GOING TO DO THIS WORK AND HOPE TO FIND THIS THING?]

2 Motivation

The motivation for developing these techniques was the on-going development of a personal automated 'Senior Companion'[5] that learns about its (senior citizen) owner's life story, needs, and preferences through free-ranging natural language dialogue with him/her, all the while offering him/her appropriate kinds of help with IT tasks, suitable entertainment, and companionable conversation. One of the many challenges in designing a free-ranging dialogue system is enabling the system to keep track of and work with the current topic of conversation: to know when a topic has been exhausted; to know how to recognise a change of topic; to be able to initiate or recognise a return to a topic that was abandoned earlier; to know when to initiate a change of topic, and what to change it to. Our aim has been to attempt to gain insight into these areas by using automated methods to spot patterns of topic change mixed with functional structure across a collection of transcribed spoken dialogues. Our goal was essentially that through machine learning methods our dialogue corpora would give up the secrets of dialogue structure, which we could then incorporate in the implementation of the Senior Companion.

[1] University of ????, COUNTRY1, email: NAME1@SERVER1
[2] University of ????, COUNTRY2, email: NAME2@SERVER2
[3] University of ????, COUNTRY3, email: NAME3@SERVER3
[4] University of ????, COUNTRY4, email: NAME4@SERVER4
[5] Being developed under *Companions*, European Commission Sixth Framework Programme Information Society Technologies Integrated Project IST-34434 (http://www.companions-project.org/).

3 Background and previous work

The linguistic ideas that underpin dialogue act ('DA') analysis are not new. [12] introduces 'speech acts' as a fundamental concept of linguistic pragmatics, analysing, for example, what it means to make a statement or ask a question. All major dialogue theories treat dialogue acts as, to some extent, a central notion, *e.g.*, in [6] or [1]. The conceptual granularity of the dialogue act labels used varies considerably between alternative analyses, often being driven by demands specific to some application or domain. However, in 1998 the Discourse Resource Initiative finalised a task-independent set of dialogue acts, called DAMSL (Dialogue Act Markup in Several Layers), for use across different domains [4]. DAMSL has been used to mark up several dialogue corpora, such as TRAINS [5], and SWITCHBOARD [8], and was the set from which the smaller set of dialogue acts used in this research was derived.

With regard to the automatic induction of dialogue action frameworks, it is clear that there are structures within dialogue that occur at levels higher than individual utterances or speaker turns. For example, [3] describe 'dialogue games' as exchanges between speakers that fulfil some limited goal, embodying our expectation of natural patterns within dialogue, *e.g.*, that a question usually precedes an answer, and a request usually precedes an acceptance or a refusal. Games consist of initiations that set up a discourse observation, and are then usually followed by responses that fulfil those expectations. Our hope was that larger and more complex structures of this sort would be identifiable through machine learning, and would give us the insight we needed into working with topic change. Evidence supporting this view can be found in [11], who achieved a 12% improvement in dialogue act tagging accuracy on the MAPTASK corpus by exploiting information of manually assigned game boundaries and types.

4 Methodology

4.1 Companions dialogue corpus collection

The automatic induction of dialogue action frameworks ('DAFs') was carried out using a spoken dialogue corpus that we collected under Companions. The corpus contains transcribed one-to-one spoken dialogues in which two participants (*A* and *B*) discussed the family photographs of one of the participants (*B*). The general aim of all these conversations was for *A* to gradually get to know *B*'s life story by means of perusing *B*'s family photos together with *A*, and discussing what (principally who) the photographs showed. This particular domain for the corpus collection was carefully chosen to provide suitable empirical data that would help with the development of the

Senior Companion.[6]

A subset of the Companions dialogues was obtained by using the 'Wizard of Oz' ('Wooz') method, in which *B* had a conversation with a computer system (*A*), but the system was actually being operated remotely by a human being. *B* was in fact aware that *A* was a human pretending to be a computer, so no deception was being practised. *A* was also observing some restrictions in his/her conversational abilities, and attempting in some carefully defined ways to converse just as it is expected the mimicked computer system will eventually converse, once it is finished. The point of using this method was to try to lead *B* into behaving more as he/she would behave if *A* really were a conversationally compromised computer system, and thus to elicit utterances from him/her that are supposedly more like the utterances that our real computer system will eventually elicit. [DGF: I AM MAKING THIS UP, IF YOU HADN'T GUESSED...]

Using the Wooz method assumes that people will talk differently to a computer that is manifestly 'worse at conversation' than to a human being. But this is an empirical question whose answer at least partially lies in how good or bad the particular computer is at pretending to be human. So in using Wooz, there is an implicit assumption that we have correctly guessed the conversational prowess of our eventual system, which, of course, we cannot have done accurately, and so the data we have collected by the Wooz method cannot be trusted. For this reason, we have not restricted the corpus we use for the automatic induction of dialogue action frameworks to the Wooz dialogues, but we have also included dialogues in which *A* presented him/herself as a human, and an openly human-to-human dialogue about *B*'s photographs was conducted.

The current Companions corpus of Wooz and human-human dialogues which was used to derive the results discussed in this paper contains [DGF, GET THIS NUMBER OF] turns and [DGF, GET THIS NUMBER OF] words. As the Companions project progresses, more dialogues are being collected for analysis, principally by recording conversations between users and the system during evaluation sessions.

4.2 Automatic dialogue act tagging

The first step in inducing dialogue action frameworks from the Companions corpus was to tag the corpus with dialogue acts, assigning one act per utterance. The set of dialogue acts we used for tagging contained [DGF, GET THE NUMBER OF] acts. These were derived from the domain-independent dialogue mark-up structure called DAMSL (see section 3). Some groups of DAMSL tags were conflated into a single DA tag, and others omitted altogether, on the grounds that, for the immediate future, we are not aiming for our system to be able to respond appropriately to *rhetorical questions*, for example, to *self-talk*, or *3rd-party talk* (all of these are DAMSL DA tags), and so there is no need for our system to be able to recognise those acts for now. While we are in the early stages of developing the system, it is important that we have high accuracy rates of dialogue act tagging, which are guaranteed with a smaller set of DA tags. Equally, it is important that we focus on getting the simpler conversational exchanges right first, and not try to run before we can walk.

The DA tagging was carried out using an n-best approach, which we have already shown can achieve high levels of accuracy [13]. We also used dialogue context to determine between some targeted n-best selection, where the perplexity of choices was lowered by using selective n-best tagging—assigning multiple categories only when we were unsure of our 1-best approach.

[6] One of the aims for the Senior Companion is for the Companion to use discussions of family photos to build a family tree and life narrative for its owner that can be used for pleasure by the owner, used by other family members, and passed on for posterity.

4.3 Automatically deciding what each utterance is *about*

[DGF: I'VE PUT THIS IN HERE SO THAT YOU CAN SEE WHAT MY THOUGHTS WERE, BUT I FULLY EXPECT YOU TO WANT TO DO SOMETHING COMPLETELY DIFFERENT.]

The next step was to automatically derive and attach to each utterance what the utterance was *about*—what the topic of the utterance was. We first carried out a frequency count of nouns across the corpus. Then we threw away the nouns that only occurred once. For every turn in the corpus, we then selected each remaining 'commonly occurring' noun, and, using WordNet, identified its immediate hypernym (the semantic mother node class) [CAN WE GET THIS FROM WORDNET?], and substituted the hypernym for its hyponym. We also substituted a 'person:NAME' marker for every occurrence of a person name; substituted a 'pronoun:PRON' marker for every third person pronoun (except *it*, which is often used as a dummy); and substituted a 'location:NAME' marker for every occurrence of a location name. In this way were were able to derive an *about set* for each utterance, containing the topics that that utterance was *about*. We took particular care with person names and location names, because these are common topics when discussing one's personal photographs. We also took particular note of the use of third-person pronouns, on the gross assumption that if such a pronoun is being used, the topic of the conversation has probably not changed, although we realise that this is greatly oversimplifying the phenomenon of anaphor resolution across dialogue.

4.4 Automatic dialogue action framework induction

[DGF: THIS GETS VERY EMBARRASSING FROM HERE ONWARDS, BECAUSE I HAVE ABSOLUTELY NO IDEA WHAT I'M DOING WITH RESPECT TO MACHINE LEARNING. PLEASE DELETE IT ALL AND WRITE SOMETHING MORE SENSIBLE, IF YOU HAVEN'T ALREADY DELETED THE WHOLE LOT, INCLUDING THIS COMMENT!]

For each utterance in the corpus, we now had an assigned dialogue act and an *about set* of topics that the utterance was *about*. Next we took sequences of the *about sets* in the order in which they occur in the corpus, and analysed how they change over time (turn number). From this analysis we derived an average number (*N*) of turns throughout which an *about set* item appeared at regular close intervals. Then, to decide whether the user had changed topic, and what he/she had changed topic to, we reasoned as follows. First we assumed the user had (probably) not changed the topic of the conversation if:

- the *about set* for the current user turn was empty.
- the *about set* for the current user turn contained one or more third person pronoun marker.
- the *about set* for the current user turn contained an item that also appeared in an earlier turn (user turn or system turn) which was not more than *N* (see above) turns earlier.

Otherwise, we assumed the user (probably) had changed topic. What had he/she changed topic to? The things that were contained

in the current *about set* that were absent from the previous *N about sets*.

Now we had patterns of topic change across the corpus, which we annotated with a mark-up, and over which mark-up we laid the dialogue act tags. [DGF: THIS IS SO BAD!!! I CAN'T BRING MYSELF TO WRITE ANY MORE. I'M SORRY. I THINK I PROBABLY SHOULDN'T EVEN BY TRYING TO DO THIS.]

4.5 Results

[DGF: Aaaarrrgggghhhh!!!!!!!!!!!!!!!! I GUESS IT'S BETTER TO ACTUALLY DO THE WORK BEFORE WE SAY WHAT OUR RESULTS ARE...]

5 CONCLUSION

[DGF: DITTO]

ACKNOWLEDGEMENTS

[DGF: SOMEONE ACKNOWLEDGEMENTS]

REFERENCES

[1] J. F. Allen, B. Miller, E. Ringger, and T. Sikorski. A robust system for natural spoken dialogue, 1996. In *Proc. 34th Annual Meeting of the Association for Computational Linguistics (ACL 96)*, Santa Cruz, California.

[2] R. Barzilay and M. Elhadad. Using lexical chains for text summarization, 1999. In I. Mani and M. T. Maybury (eds), *Advances in Automatic Text Summarization*, Chapter 10, MIT Press.

[3] J. Carletta, A. Isard, S. Isard, J. C. Kowtko, G. Doherty-Sneddon, and A. H. Anderson. The reliability of a dialogue structure coding scheme, 1997. *Computational Linguistics* 23(1): 13-31.

[4] M. Core, M. Ishizaki, J. Moore, C. Nakatani, N. Reithinger, D. Traum, and S. Tutiya. *The report of the third workshop of the discourse resource initiative*, 1999.

[5] M. G. Core and J. Allen. Coding dialogs with the DAMSL annotation scheme, 1997. In *AAAI Fall Symposium on Communicative Action in Humans and Machines*, MIT, Cambridge, MA.

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

[7] M. Hearst. TextTiling: Segmenting text into multi-paragraph subtopic passages, 1997. *Computational Linguistics* 23(1): 33–64.

[8] D. Jurafsy, R. Bates, N. Coccaro, R. Martin, M. Meteer, K. Ries, E. Shriberg, A. Stolcke, P. Taylor, and C. Van Ess-Dykema. Switchboad discourse language modeling project final report., 1998. *Research Note 30*, Center for Language and Speech Processing, Johns Hopkins University, Baltimore.

[9] H. Kozima. Text segmentation based on similarity between words., 1993. In *31st Annual Meeting of the Association for Computational Linguists (Student Session)*, pp. 286–288.

[10] J. Morris and G. Hirst. Lexical cohesion computed by thesaural relations as an indicator of the structure of text, 1991. *Computational Linguistics* 17:(1) 21–48.

[11] M. Poesio and A. Mikheev. The predictive power of game structure in dialogue act recognition: Experimental results using maximum entropy estimation, 1998. In *Proc. International Conference on Speech and Language Processing (ICSLP-98)*, Australia. Reithinger: Norbert and Martin.

[12] J. R. Searle. *Speech acts: An essay in the philosophy of language*, 1969. Cambridge: Cambridge University Press.

[13] N. Webb, M. Hepple, and Y. Wilks. Dialogue act classification based on intra-utterance features, 2005. In *Proc. AAAI workshop on spoken language understanding*, Pittsburgh.