Using NLG to Help Language-Impaired Users Tell Stories and Participate in Social Dialogues

Ehud Reiter, Ross Turner
University of Aberdeen
Aberdeen, UK
e.reiter@abdn.ac.uk
csc272@abdn.ac.uk

**Norman Alm, Rolf Black,
Martin Dempster, Annalu Waller**
University of Dundee
Dundee, UK
{nalm,rolfblack,martindempster,
awaller}@computing.dundee.ac.uk

Abstract

Augmentative and Alternative Communication (AAC) systems are communication aids for people who cannot speak because of motor or cognitive impairments. We are developing AAC systems where users select information they wish to communicate, and this is expressed using an NLG system. We believe this model will work well in contexts where AAC users wish to go beyond simply making requests or answering questions, and have more complex communicative goals such as story-telling and social interaction.

1 Introduction

Many people have difficulty in communicating linguistically because of cognitive or motor impairments. Such people typically use communication aids to help them interact with other people. Such communication aids range from simple tools that do not involve computers, such as picture cards, to complex software systems that attempt to "speak" for the impaired user.

From a technological perspective, even the most complex communication aids have typically been based on fixed (canned) texts or simple fill-in-the-blank templates; essentially the user selects a text or template from a set of possible utterances, and the system utters it. We believe that while this may be adequate if the user is simply making a request (e.g., *please give me a drink*) or answering a question (e.g., *I live at home*), it is not adequate if the user has a more complex communicative goal, such as engaging in social interaction, or telling a story.

We are exploring the idea of supporting such interactions by building a system which uses external data and/or knowledge sources, plus do-main and conversational models, to dynamically suggest possible *messages* (event, facts, or opinions, represented as ontology instances) which are appropriate to the conversation. The user selects the specific message which he wishes the system to speak, and possibly adds simple annotations (e.g., *I like this*) or otherwise edits the message. The system then creates an appropriate linguistic utterance from the selected message, taking into consideration contextual factors.

In this paper we describe two projects on which we are working within this framework. The goal of the first project is to help non-speaking children tell stories about their day at school to their parents; the goal of the second project is to help non-speaking adults engage in social conversation.

2 Background

2.1 Augmentative and alternative communication

Augmentative and alternative communication (AAC) is a term that describes a variety of methods of communication for non-speaking people which can supplement or replace speech. The term covers techniques which require no equipment, such as sign language and cards with images; and also more technologically complex systems which use speech synthesis and a variety of strategies to create utterances.

The most flexible AAC systems allow users to specify arbitrary words, but communication rates are extremely low, averaging 2-10 words per minute. This is because many AAC users interact slowly with computers because of their impairments. For example, some of the children we work with cannot use their hands, so they use scanning interfaces with head switches. In other words, the computer displays a number of op-

tions to them, and then scans through these, briefly highlighting each option. When the desired option is highlighted, the child selects it by pressing a switch with her head. This is adequate for communicating basic needs (such as hunger or thirst); the computer can display a menu of possible needs, and the child can select one of the items. But creating arbitrary messages with such an interface is extremely slow, even if word prediction is used; and in general such interfaces do not well support complex social interactions such as story telling (Waller, 2006).

A number of research projects in AAC have developed prototype systems which attempt to facilitate this type of human-human interaction. At their most basic, these systems provide users with a library of fixed "conversational moves" which can be selected and uttered. These moves are based on models of the usual shape and content of conversational encounters (Todman & Alm, 2003), and for example include standard conversational openings and closings, such as *Hello* and *How are you*. They also include back-channel communication such as *Uh-huh*, *Great!,* and *Sorry, can you repeat that*.

It would be very useful to go beyond standard openings, closings, and backchannel messages, and allow the user to select utterances which were relevant to the particular communicative context and goals. Dye et al (1998) developed a system based on scripts of common interactions (Schank & Abelson, 1977). For example, a user could activate the MakeAnAppointment script, and then could select utterances relevant to this script, such as *I would like to make an appointment to see the doctor*. As the interaction progressed, the system would update the selections offered to the user based on the current stage of the script; for example during time negotiation a possible utterance would be *I would like to see him next week*. This system proved effective in trials, but needed a large number of scripts to be generally effective. Users could author their own texts, which were added to the scripts, but this was time-consuming and had to be done in advance of the conversation.

Another goal of AAC is to help users narrate stories. Narrative and storytelling play a very important part in the communicative repertoire of all speakers (Schank, 1990). In particular, the ability to draw on episodes from one's life history in current conversation is vital to maintaining a full impression of one's personality in dealing with others (Polkinghorne, 1991). Story telling tools for AAC users have been developed, which include ways to introduce a story, tell it at the pace required (with diversions) and give feedback to comments from listeners (Waller, 2006); but again these tools are based on a library of fixed texts and templates.

2.2   NLG and AAC

Natural language generation (NLG) systems generate texts in English and other human languages from non-linguistic input (Reiter and Dale, 2000). In their review of NLP and AAC, Newell, Langer, and Hickey (1998) suggest that NLG could be used to generate complete utterances from the limited input that AAC users are able to provide. For example, the Compansion project (McCoy, Pennington, Badman 1998) used NLP and NLG techniques to expand telegraphic user input, such as *Mary go store?,* into complete utterances, such as *Did Mary go to the store*? Netzer and Elhadad (2006) allowed users to author utterances in the symbolic language BLISS, and used NLG to translate this to English and Hebrew texts.

In recent years there has been growing interest in *data-to-text* NLG systems (Reiter, 2007); these systems generate texts based on sensor and other numerical data, supplemented with ontologies that specify domain knowledge. In principle, it seems that data-to-text techniques should allow NLG systems to provide more assistance than the syntactic help provided by Compansion. For example, if the user wanted to talk about a recent football (soccer) match, a data-to-text system could get actual data about the match from the web, and generate potential utterances from this data, such as *Arsenal beat Chelsea 2-1* and *Van Persie scored two goals*; the user could then select one of these to utter.

In addition to helping users interact with other people, NLG techniques can also be used to educate and encourage children with disabilities. The STANDUP system (Manurung, Ritchie et al., 2008), for example, used NLG and computational humour techniques to allow children who use AAC devices to generate novel punning jokes. This provided the children with successful experiences of controlling language, gave them an opportunity to play with language and explore new vocabulary (Waller et al., in press). In a small study with nine children with cerebral palsy, the children used their regular AAC tools more and also performed better on a test measuring linguistic abilities after they used STANDUP for ten weeks.

3   Our Architecture

Our goal is help AAC users engage in complex social interaction by using NLG and data-to-text technology to create potential utterances and conversational contributions for the users. The general architecture is shown in Figure 1, and Sections 4 and 5 describe two systems based on this architecture.
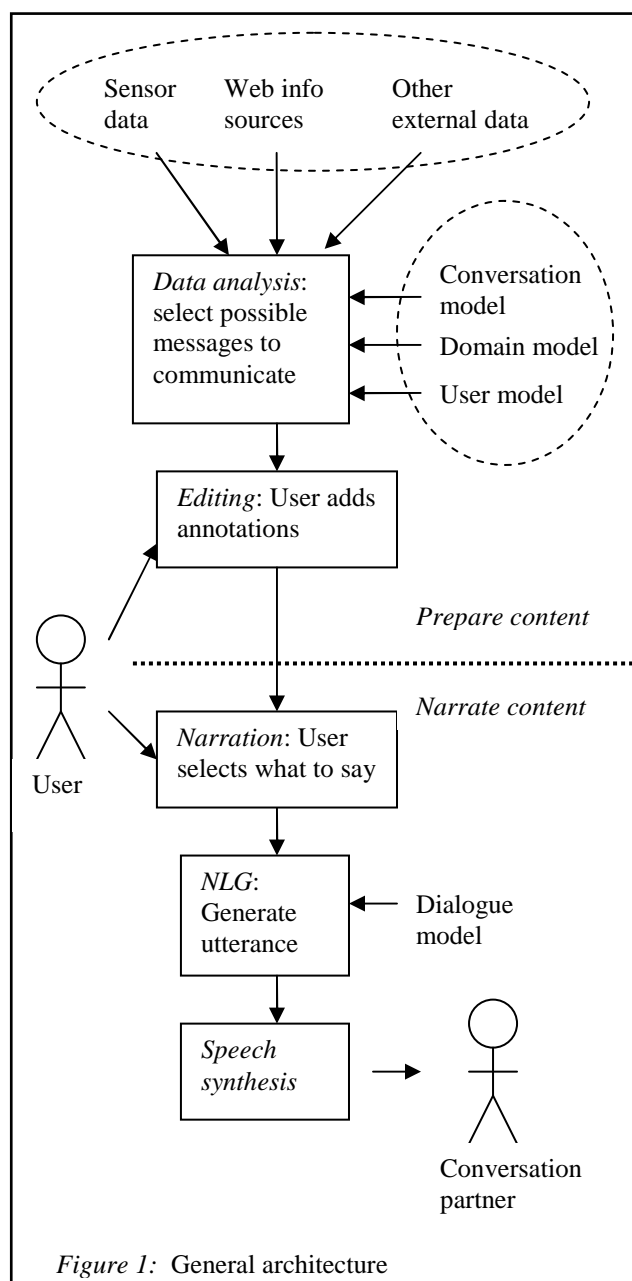

Figure 1: General architecture

The system has the following components:

Data analysis: read in data, from sensors, web information sources, databases, and so forth. This module analyses this data and identifies messages (in the sense of Reiter and Dale (2000)) that the user is likely to want to communicate; this analysis is partially based on domain, conversation, and user models, which may be represented as ontologies.

Editing: allow the user to edit the messages. Editing ranges from adding simple annotations to specify opinions (e.g., add BAD to *Arsenal beat Chelsea 2-1* if the user is a Chelsea fan), to using an on-screen keyboard to type free-text comments. Users can also delete messages, specify which messages they are most likely to want to utter, and create new messages. Editing is done before the actual conversation, so the user does not have to do this under time pressure. The amount of editing which can be done partially depends on the extent of the user's disabilities.

Narration: allows the user to select messages, and perhaps conversational moves (e.g., *Hello*), in an actual conversational context. Editing is possible, but is limited by the need to keep the conversation flowing.

NLG and Speech Synthesis: Generates actual utterances from the selected messages, taking into account linguistic context, especially a dialogue model.

4   Narrative for Children: *How was School Today*

The goal of the *How was School Today* project is to enable non-speaking children with major motor disabilities but reasonable cognitive skills to tell a story about what they did at school during the day. The particular children we are working with have cerebral palsy, and use wheelchairs. A few of them can use touch screens, but most of them use a head switch and scanning interface, as described above. By 'story', we mean something similar to Labov's (1972) conversational narrative, i.e., a series of linked real-world events which are unusual or otherwise interesting, possibly annotated with information about the child's feelings, which can be narrated orally. We are not expecting stories in the literary sense, with character development and complex plots.

The motivation of the project is to provide the children with successful narrative experience. Typically developing children develop narrative skills from an early age with adults scaffolding conversations to elicit narrative, e.g. *"What did you do at school today?"* (Bruner, 1975). As the child's vocabulary and language competence develops, scaffolding is reduced. This progression is seldom seen in children with complex communication needs – they respond to closed questions but seldom take control of conversa-

tion (von Tetzchner and Grove, 2003). Many children who use AAC have very limited narrative skills (Soto et al, 2006). Research has shown that providing children who use AAC with successful narrative experiences by providing full narrative text can help the development of written and spoken narrative skills (Waller, 2008).

The system follows the architecture described above. Input data comes from RFID sensors that track where the child went during the day; an RFID reader is mounted on the child's wheelchair, and RFID tags are placed around the school, especially in doorways so we can monitor children entering and leaving rooms. Teachers have also been given RFID swipe cards which they can swipe against a reader, to record that they are interacting with the child; this is more robust than attempting to infer interaction automatically by tracking teachers' position. Teachers can also record interactions with objects (toys, musical instruments, etc), by using special swipe cards associated with these objects. Last but not least, teachers can record spoken messages about what happened during the day. An example of how the child's wheelchair is set up is shown in Figure 2.
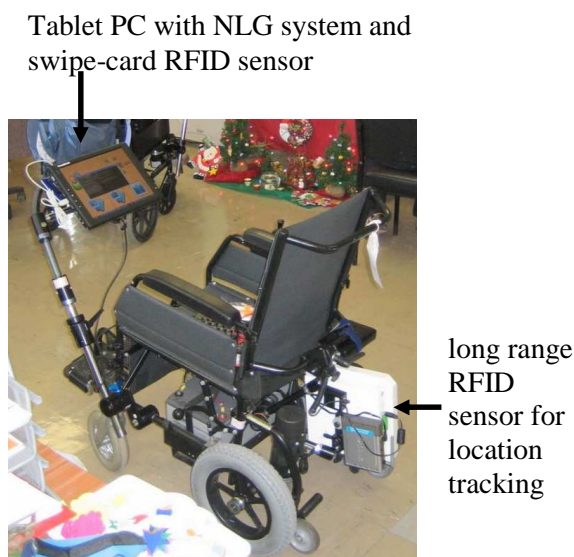

Figure 2: System configuration

The data analysis module combines sensor-derived location and interaction data with a timetable which records what the child was expected to do during the day, and a domain knowledge base which includes information about typical activities (e.g., if the child's location is SwimmingPool, the child's activity is probably Swimming). From this it creates a series of events (each of which contain a number of messages) which describe the child's lessons and activities, including divergences from what is expected in the timetable. Several messages may be associated with an event. The data analysis module also infers which events and messages it believes are most interesting to the child; this is partially based on heuristics about what children are interested in (e.g., swimming is more interesting than lunch), and partially based on the general principle that unexpected things (divergences from the timetable) are more interesting than expected things. No more than five events are flagged as interesting, and only these events are shown in the editing interface.

The editing interface allows children to remove events they do not want to talk about (perhaps for privacy reasons) from the list of interesting events. It also allows children to add messages that express simple opinions about events; i.e., *I liked it* or *I didn't like it*. The interface is designed to be used with a scanning interface, and is based on symbols that represent events, annotations, etc.

The narration interface, shown in Figure 3, is similar to the editing interface. It allows children to choose a specific event to communicate, which must be one of the ones they selected during the editing phase. Children are encouraged to tell events in temporal order (this is one of the narration skills we are trying to teach), but this is not mandated, and they can deviate from temporal order if they wish.
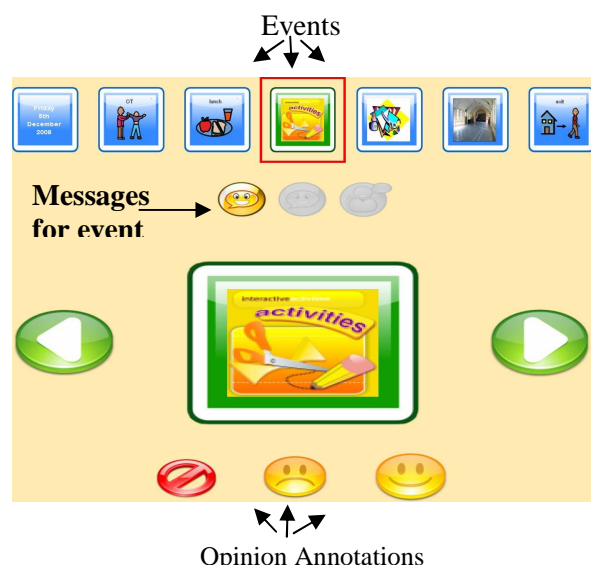

Figure 3: Narration Interface

The NLG system generates actual texts from the events selected by the children. Most of this

is fairly simple, since the system deliberately uses simple "child-like" language (Section 6). However, the system does need to make some decisions based on discourse context, including choosing appropriate referring expressions (especially pronouns), and temporal expressions (especially when children deviate from pure temporal order).

4.1 Example

For example, assume that the timetable specifies the following information

Time	Activity	Location	Teacher
……	……	……	……
13.20 -14	Arts and Crafts	CL_SEC2	Mrs Smith
14 -14.40	Physiotherapy	PHYSIO1	Mrs Jones
……	……	……	……

Assume that the sensors then recorded the following information

Event 1
    Location: CL_SEC2
    Time: 13:23:00.0 - 14:07:00.0
    Interactions: Mrs. Smith, Rolf, Ross

Event 2
    Location: HALL
    Time: 14:10:00.0 – 14:39:00.0
    Interactions: none

The data analysis module associates Event 1 with the Arts and Crafts timetable entry, since the location is right, the timetabled teacher is present, and the times approximately match. From this two messages are produced: one corresponding to *I had Arts and Crafts this afternoon with Mrs. Smith* (the core activity description), and the other corresponding to *Rolf and Ross were there* (additional information about people not timetabled to be there). The child can add opinions using the editing interface; for example, if he added a positive annotation to the event, this would become an additional message corresponding to *It was great*.

For Event 2, the data analysis module notes that it does not match a timetabled event. The timetable indicates the child should be at Physiotherapy after Art and Crafts; however, the sensor information indicates they were in the hall. The system generates a single message corresponding to *Then I went to the Hall instead of Physiotherapy* to describe this event. If the child added a negative annotation to this message, this would become an additional message expressed as *I didn't like it.*

4.2 Evaluation

We conducted an initial evaluation of the How was School Today system in January, 2009. Two children used the system for four days: Julie, age 11, who had good cognitive skills but was non-verbal because of severe motor impairments; and Jessica, age 13, who had less severe motor impairments but who had some cognitive and memory impairments (these are not the childrens' real names). Julie used the system as a communication and interaction aid, as described above; Jessica used the system partially as a memory aid. The evaluation was primarily qualitative: we observed how Julie and Jessica used the system, and interviewed their teachers, speech therapists, care assistants, and Julie's mother (Jessica's parents were not available).

The system worked very well for Julie; she learned it quickly, and was able to use it to have real conversations about her day with adults, almost for the first time in her life. This validated our vision that our technology could help AAC users engage in real interaction, and go beyond simple question answering and communication of basic needs. The system also worked reasonably well as a memory aid for Jessica, but she had a harder time using it, perhaps because of her cognitive impairments.

Staff and Julie's mother were very supportive and pleased with the system. They had suggestions for improving the system, including a wider range of annotations; more phrases about the conversation itself, such as *Guess what happened at school today*; and allowing children to request teenager language (e.g., *really cool*).

From a technical perspective, the system worked well overall. School staff were happy to use the swipe cards, which worked well. There were some problems with the location sensors, we need better techniques for distinguishing real readings from noise. A surprising amount of effort was needed to enter up-to-date knowledge (e.g., daily lunch menus), this would need to be addressed if the system was used for a period of months as opposed to days.

5 Social Conversation for Adults

In our second project, we want to build a tool to help adults with cerebral palsy engage in social conversation about a football match, movie, weather, and so forth. Many people with severe disabilities have great difficulty developing new interpersonal relationships, and indeed report that forming new relationships and taking part in new

activities are major priorities in their lives (Datillo et al., 2007). Supporting these goals through the development of appropriate technologies is important as it could lead to improved social outcomes.

This project builds on the TALK system (Todman and Alm, 2003), which helped AAC users engage in active social conversation. TALK partially overcame the problem of low communication rate by requiring users to pre-author their conversational material ahead of time, so that when it was needed it could simply be selected and output. TALK also used insights from Conversation Analysis (Sacks, 1995) to provide appropriate functionality in the system for social conversation. For example, it supported opening and closing statements, stepwise topic change, and the use of quick-fire utterances to provide fast, idiomatic responses to commonly encountered situations. This approach led to more dynamic AAC-facilitated interactions with higher communication rates, and had a positive impact on the perceived communicative competence of the user (Todman, Alm et al., 2007).

TALK requires the user to spend a substantial amount of time pre-authoring material; this is perhaps its greatest weakness. Our idea is to reduce the amount of pre-authoring needed, by using the architecture shown in Fig 1, where much of the material is automatically created from data sources, ontologies, etc, and the user's role is largely to edit and annotate this material, not to create it from scratch.

We developed an initial prototype system to demonstrate this concept in the domain of football results (Dempster, 2008). We are now working on another prototype, whose goal is to support social conversations about movies, music, television shows, etc (which is a much broader domain than football). We have created an ontology which can describe *events* such as watching a film, listening to a music track, or reading a book. Each 'event' has both temporal and spatial properties which allow descriptions to be produced about where and when an event took place, and other particulars relating to that particular class of event. For example, if the user listened to a radio show, we record the name of the show, the presenter and the station it was broadcast on. Ultimately we plan to obtain information about movies, music tracks, etc from web-based databases such as IMDB (movies) and last.fm (music).

Of course, databases such as IMDB do not contain information such as what the user thought of the movie, or who he saw it with. Hence we will allow users to add annotations with such information. Some of these annotations will be entered via a structured tool, such as a calendar interface that allows users to specify when they watched or listened to something. We would like to use NaturalOWL (Galanis and Androutsopoulos, 2007) as the NLG component of the system; it is well suited to describing objects, and is intended to be integrated with an ontology. As with the How Was School Today project, some of the main low-level NLG challenges are choosing appropriate referring expressions and temporal references, based on the current discourse context. Speech output is done using Cereproc (Aylett and Pidcock, 2007).

An example of our current narration interface is shown in Figure 4. In the editing interface, the user has specified that he went to a concert at 8pm on Thursday, and that he rated it 8 out of 10. The narration interface gives the user a choice of a number of messages based on this information, together with some standard messages such as *Thanks* and *Agree.*
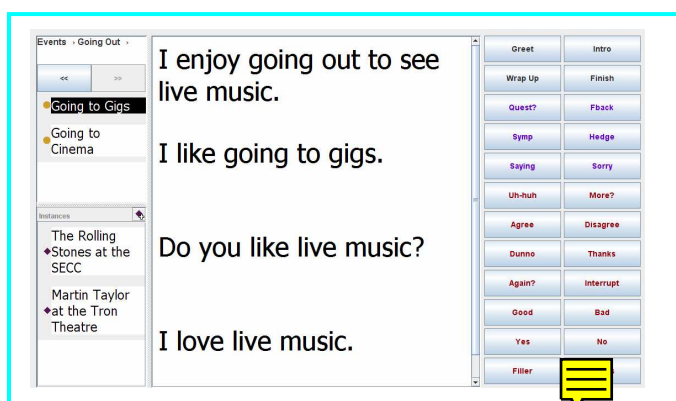
Note that unlike the How Was School Today project, in this project we do not attempt to infer event information from sensors, but we allow (and expect) the user to enter much more information at the editing stage. We could in principle use sensors to pick up some information, such as the fact that the user was in the cinema from 12 to 2PM on Tuesday, but this is not the research focus of this project.

We plan to evaluate the system using groups of both disabled and non-disabled users. This has been shown in the past to be an effective approach for the evaluation of prototype AAC systems (Higginbotham, 1995). Initially pairs of non-disabled participants will be asked to produce short conversations with one person using the prototype and the other conversing normally. Quantitative measures of the communication rate

will be taken as well as more qualitative observations relating to the usability of the system. After this evaluation we will improve the system based on our findings, and then conduct a final evaluation with a small group of AAC users.

6 Discussion: Challenges for NLG

From an NLG perspective, generating AAC texts of the sort we describe here presents different challenges from many other NLG applications.

First of all, realization and even microplanning are probably not difficult, because in this context the AAC system should generate short simple sentences if possible. This is because the system is speaking "for" someone with limited or developing linguistic abilities, and it should try to produce something similar to what the user would say himself if he or she had the time to explicitly write a text using an on-screen keyboard.

To take a concrete example, we had originally considered using past-perfect tense (a fairly complex linguistic construct) in the How was School project, when the narrative jumped to an earlier point in time. For example *I ate lunch at 12. I had gone swimming at 11*. But it was clear from corpora of child-written texts that these children never used perfect tenses, so instead we opted for *I ate lunch at 12. I went swimming at 11*. This is less linguistically polished, but much more in line with what the children might actually produce.

Given this desire for linguistic simplicity, realisation is very simple, as is lexical choice (use simple words) and aggregation (keep sentences short). The main microplanning challenges relate to discourse coherence, in particular referring expressions and temporal descriptions.

On the other hand, there are major challenges in document planning. In particular, in the How Was School project, we want the output to be a proper narrative, in the sense of Labov (1972). That is, not just a list of facts and events, but a structure with a beginning and end, and with explanatory and other links between components (e.g., *I had math in the afternoon because we went swimming in the morning*, if the child normally has math in the morning). We also wanted the narrative to be interesting and hold the interest of the person the child is communicating with. As pointed out by Reiter et al (2008), current NLG systems do not do a good job of generating narratives.

Similarly, in the Social Conversations project we want the system to generate a social dialogue, not just a list of facts about movies and songs.

Little previous research has been done on generating social (as opposed to task-oriented) dialogues. One exception is the NECA Socialite system (van Deemter et al, 2008), but this focused on techniques for expressing affect, not on high-level conversational structure.

For both stories and social conversations, it would be extremely useful to be able to monitor what the conversational partner is saying. This is something we hope to investigate in the future. As most AAC users interact with a small number of conversational partners, it may be feasible to use a speech dictation system to detect at least some of what the conversational partner says.

Last but not least, a major challenge implicit in our systems and indeed in the general architecture is letting users control the NLG system. Our systems are intended to be speaking aids, ideally they should produce the same utterances as the user would if he was able to talk. This means that users must be able to control the systems, so that it does what they want it to do, in terms of both content and expression. To the best of our knowledge, little is known about how users can best control an NLG system.

7 Conclusion

Many people are in the unfortunate position of not being able to speak or type, due to cognitive and/or motor impairments. Current AAC tools allow such people to engage in simple needs-based communication, but they do not provide good support for richer use of language, such as story-telling and social conversation. We are trying to develop more sophisticated AAC tools which support such interactions, by using external data and knowledge sources to produce candidate messages, which can be expressed using NLG and speech synthesis technology. Our work is still at an early stage, but we believe that it has the potential to help AAC users engage in richer interactions with other people.

References

Aylett, M. and C. Pidcock (2007). The CereVoice Characterful Speech Synthesiser SDK. *Proceedings of Proceedings of the 7th International Conference on Intelligent Virtual Agents*, pages 413-414.

Bruner, J. (1975). From communication to language: A psychological perspective. *Cognition* **3**: 255-289.

Datillo, J., G. Estrella, L. Estrella, J. Light, D. McNaughton and M. Seabury (2007). "I have chosen to live life abundantly": Perceptions of leisure by adults who use Augmentative and Alternative Communication. *Augmentative & Alternative Communication* **24**(1): 16-28.

van Deemter, K., B Krenn, P Piwek, M Klesen, M Schröder and S Baumann. Fully generated scripted dialogue for embodied agents. *Artificial Intelligence* **172**: 1219–1244.

Dempster, M. (2008). Using natural language generation to encourage effective communication in non-speaking people. *Proceedings of Young Researchers Consortium, ICCHP'08*.

Dye, R., N. Alm, J. Arnott, G. Harper, and A. Morrison (1998). A script-based AAC system for transactional interaction. *Natural Language Engineering*, *4*(1), 57-71.

Galanis, D. and I. Androutsopoulos (2007). Generating Multilingual Descriptions from Linguistically Annotated OWL Ontologies: the NaturalOWL System. *Proceedings of ENLG 2007*.

Higginbotham, D. J. (1995). Use of nondisabled subjects in AAC Research : Confessions of a research infidel. *Augmentative and Alternative Communication* **11**(1): 2-5.

Labov, W (1972). *Language in the Inner City*. University of Pennsylvania Press.

Manurung, R., G. Ritchie, H. Pain, A. Waller, D. O'Mara and R. Black (2008). The Construction of a Pun Generator for Language Skills Development. *Applied Artificial Intelligence* **22**(9): 841 – 869.

McCoy, K., C. Pennington and A. Badman (1998). Compansion: From research prototype to practical integration. *Natural Language Engineering* **4**:73-95.

Netzer, Y and Elhadad, M (2006). Using Semantic Authoring for Blissymbols Communication Boards. In *Proc of HLT-2006*.

Newell, A., S. Langer and M. Hickey (1998). The role of natural language processing in alternative and augmentative communication. *Natural Language Engineering* **4**:1-16.

Polkinghorne, D. (1991). Narrative and self-concept. *Journal of Narrative and Life History*, **1**(2/3), 135-153

Reiter, E (2007). An Architecture for Data-to-Text Systems. In *Proceedings of ENLG-2007, pages* 147-155.

Reiter, E. and R. Dale (2000). *Building Natural Language Generation Systems*. Cambridge University Press.

Reiter, E, A. Gatt, F Portet, and M van der Meulen (2008). The Importance of Narrative and Other Lessons from an Evaluation of an NLG System that Summarises Clinical Data (2007). In *Proceedings of INLG-2008,* pages 97-104.

Sacks, H. (1995). *Lectures on Conversation*. G. Jefferson. Cambridge, MA, Blackwell.

Schank, R. C. (1990). *Tell me a story: A new look at real and artificial intelligence*. New York, Macmillan Publishing Co.

Schank, R., and R. Abelson (1977). *Scripts, plans, goals, and understanding*. New Jersey: Lawrence Erlbaum.

Soto, G., E. Hartmann, and D. Wilkins (2006). Exploring the Elements of Narrative that Emerge in the Interactions between an 8-Year-Old Child who uses an AAC Device and her Teacher. *Augmentative and Alternative Communication* **4**:231 – 241.

Todman, J. and N. A. Alm (2003). Modelling conversational pragmatics in communication aids. *Journal of Pragmatics* **35**: 523-538.

Todman, J., N. A. Alm, D. J. Higginbotham and P. File (2007). Whole Utterance Approaches in AAC. *Augmentative and Alternative Communication* **24**(3): 235-254.

von Tetzchner, S. and N. Grove (2003). The development of alternative language forms. In S. von Tetzchner and N. Grove (eds), *Augmentative and Alternative Communication: Developmental Issues*, pages 1-27. Wiley.

Waller, A. (2006). Communication Access to Conversational Narrative. *Topics in Language Disorders* **26**(3): 221-239.

Waller, A. (2008). Narrative-based Augmentative and Alternative Communication: From transactional to interactional conversation. *Proceedings of ISAAC 2008,* pages 149-160.

Waller, A., R. Black, D. A. O'Mara, H. Pain, G. Ritchie and R. Manurung (In Press). Evaluating the STANDUP Pun Generating Software with Children with Cerebral Palsy. *ACM Transactions on Accessible Computing*.