CROCODIAL: Crosslingual Computer-mediated Dialogue

Paul Piwek[1] and Richard Power[2]

[1] Centre for Research in Computing, Open University, Milton Keynes, UK,

Abstract. We describe a novel approach to *crosslingual dialogue* which allows for *highly accurate* communication of *semantically complex* content. The approach is introduced through an application in a B2B scenario. We are currently building a browser-based prototype for this scenario. The core technology underlying the approach is natural language generation. We also discuss how the proposed approach can complement Machine Translation-based solutions to crosslingual dialogue.

1 Introduction

> "The most pronounced impact of Internet technology is that it allows for human-to-human collaboration, negotiation and transactions, instead of the phone, fax or mail, collaboration can take place in real time using a browser and the Internet." (Harvey Seegers, CEO of Global eXchange Services. January 22, 2003 for CNET radio and ZDNet)

It should come as no surprise that companies such as Global eXchange Services (GXS), one of the leading providers of B2B (Business to Business) services, see the move from (e)mail, phone and fax to human-to-human interaction through a browser as a significant one: a browser provides the platform for integrating many value-added services into the functionality which has traditionally been provided by (e)mail, phone and fax. From the manifold of services which one can imagine, we focus on two: *crosslinguality* and *knowledge management*.

2 Transaction to Tuscany

Harry, a pensioner who is currently living in London, has decided that it is time to start enjoying the better things in life. He buys a villa in Tuscany from Count Roberto da Silva and instructs his bank to transfer his payment for the purchase to Da Silva's account.

Two weeks after issuing the instruction, Harry receives a phone call from Da Silva. He explains in agitated and broken English that the money has not yet been credited

Piwek P. and Power R. (2006).
CROCODIAL: Crosslingual Computer-mediated Dialogue.
In *Proceedings of the 3rd International Workshop on Computer Supported Activity Coordination*, pages 75-84
DOI: 10.5220/0002484400750084

to his account. Harry contacts the call centre of his bank and is connected to the local branch where he issued the instruction. They promise to find out what has happened. Next day, Harry contacts his local branch again, of course, via the central call centre. They reassure him that they will soon get back to him.

A few calls later, the problem is finally resolved. It transpires that the money does not show up on Da Silva's euro account because it has been routed to a special account for pounds sterling.

Probably quite a few of us have experienced similar mishaps. One of the leading clearinghouses for banks (BACS), claims on its website that they alone deal with 4.5 billion financial transactions a year.[3] Complications like the aforementioned have two unfortunate consequences for the bank. Firstly, it frustrates their customers. Secondly, the bank wastes a lot of its own time, as well as the customer's, on inefficient telephone calls. Let us imagine how the same transaction could have been dealt with in a better possible world.

Harry again receives a call from Da Silva and contacts his local branch. This time the bank tells him that he will be informed about the whereabouts of the money within two days. He is asked whether he wants to receive the information by telephone or email. Harry prefers email. Next, an employee of Harry's bank connects with her browser to the server of the bank's clearinghouse and requests information on the transfer. It transpires that the money has been successfully transferred to the account of Da Silva.

The employee selects the option to engage in a crosslingual dialogue with an employee of the Italian paying-bank (the bank which received the money). After a brief interaction, it becomes clear that the money has been placed in Da Silva's pound sterling account. The content of the conversation is kept on record and used to automatically generate a summary in English for Harry. This is delivered to him by email.

The scenario exemplifies a wider problem that faces many businesses that operate globally: language barriers and the distribution of tasks among a multitude of partners – e.g., the collecting and paying banks with local branches and head offices and the clearinghouse – hamper smooth global interactions and reduce customer satisfaction. The core problem is that information needs to be interchanged reliably in different languages and has to be readily available for different *purposes* (company internal, B2B, B2C). Looking beyond the banking industry, there are many further scenarios along the same lines. Think, for instance, of ship-to-shore communications, where crews are made up of many nationalities, and border crossing communications between medics.

3 Prospects of Machine Translation Based Solutions

Solutions based on Machine Translation (MT) present themselves as an obvious candidate for overcoming language barriers. In recent years MT has experienced a revival, partly due to the increased demand and possibilities for translation caused by the advent of the Internet. For instance, in crosslingual information retrieval, where large volumes of text need to be translated, MT has proved very useful. Here, we are, however, considering applications that involve only small quantities of information that need to be

[3] http://www.bacs.co.uk/BPSL/corporate/corporateoverview/; accessed February 3, 2006.

exchanged with *extremely high accuracy*, because either the financial stakes are high, or the situation is safety-critical.

Unfortunately, some have predicted that high accuracy translations of text/speech input are not likely to be realized in the near future (e.g., Hutchins, 1999). Even for relatively simple domains, such as travel planning, medium and extremely large scale research projects such as the Spoken Language Translator (Rayner et al., 2000) and Verbmobil[4] have, despite making substantial contributions to various areas of speech and language processing, not yet delivered systems for practical deployment. Rayner et al. (2000) estimate that spoken language translation will eventually be possible – though still challenging and only for closed domains – with a coverage of 85 to 90%. One of the few deployed crosslingual communication systems, Linguanet,[5] relies on the use of message templates together with MT technology to achieve a level of accuracy that is acceptable for a practical application (message passing between European police forces).

Apart from high levels of accuracy, what is also largely missing in existing MT systems is the representation of the semantic and discourse content of utterances. Some systems use an interlingua, i.e., a language independent representation of the content of an utterance (e.g., Lonsdale et al. (1994)). However, most do not include coreference relationships across sentence boundaries, let alone more sophisticated anaphoric relationship such as part-whole and action-actor. Representation of content is important because it enables extra services. Consider the automatic delivery of a summary to our protagonist called Harry: a formal representation of the dialogue between the bank employees would enable a summarization program to reliably determine what conclusions were reached.

4 The Role of Context in Communication

In this section, we prepare the ground for a novel solution to the problem of accurate crosslingual communication. We describe a view of communication which differs from the view that informs existing approaches to crosslingual communication. Existing approaches are typically grounded in the classical transmission model of communication: *A* wants to communicate a certain message *m* to *B*. She encodes this into a (spoken or written) natural language sentence. *B* receives the sentence and decodes it into the message *m*. According to this view, crosslingual communication from language L_1 to L_2 reduces to the task of finding a sentence in L_2 which conveys the same message as a given sentence s in L_1.

We want to draw attention to a fundamental shortcoming of the transmission model. Since the seventies, work in both linguistics and philosophy has moved towards a rather different view of communication (e.g., Isard, 1975). Whereas the classical model is static – sentences (or better, utterances) are paired with meanings – the alternative is *dynamic*: utterances change the context, and the way in which they change that context is again dependent on context.

[4] http://verbmobil.dfki.de/

[5] http://www.prolingua.co.uk/Linguanet/index.html

For our purposes, the context includes a record of the conversational content (the dialogue history) and any relevant background information. Let us illustrate how context-dependence plays a role in communication by examining the following utterance: 'Greenspan stopped decreasing the interest rates'. The verb 'stop' is said to trigger a presupposition: a constraint on the contexts in which this utterance can be produced. The informational content of this utterance can only be accepted in a context where the interlocutor is also willing to accept that *Greenspan was decreasing the interest rates*.

Presupposed information differs from what is asserted in that it is not affected by negation. In 'Greenspan did not stop decreasing the interest rates', the assertion perishes but the presupposition survives. The example illustrates that utterances do not convey neat self-contained messages. Rather, they contribute to a context which is a network of interlinked units of information (e.g., the information that Greenspan *stopped* decreasing the interest rates depends on the information that he *was* decreasing them).

The context-change view of communication suggests a new approach to crosslingual dialogue. In dialogue the interlocutors change the context by producing utterances that extend the dialogue history. Changes at the informational level are arrived at via *interpretation* of physical actions. Now imagine that interlocutors could directly edit the context at the informational level but each see the results of their actions in a representation suitable for them: no translation would be required since each interlocutor would directly operate on the underlying *content*.

Conventional WIMP (for Windows, Icons, Menus and Pointer) interfaces allow users to do something similar: if I want to get rid of a file on my computer, instead of saying 'delete file foo', I can pick up and drop the file into the recycle bin. I receive feedback on the effects of my actions via a graphical interface. The desktop that I see is rendered on the basis of an underlying model. Different desktops can be rendered from the same underlying model without translation between desktops.

5 Crosslingual Dialogue as Joint Knowledge Editing

The contents that we can transfer by means of natural language are, of course, different from the information rendered by a windows desktop. For instance, logical vocabulary such as 'not', 'or' and 'most' introduces content for which natural graphical representations do not exist.

5.1 WYSIWYM Content Editing

The WYSIWYM technology – What You See Is What You Meant (Power et al., 1998) – presents a solution to the visualization problem. Content is rendered in natural language using natural language generation technology. The basic idea underlying WYSIWYM is presented in Figure 1.

Figure 1 represents an editing cycle. Given a Knowledge Base (KB), the system generates a description of the knowledge base in the form of a *feedback text* containing *anchors* representing places where the content in the knowledge base (a formal representation of the context) can be extended. Each anchor is associated with pop-up menus, which present the possible extensions of the KB at that point. On the basis of

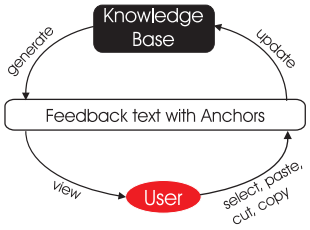

Fig. 1. The editing cycle.

the extension that the user selects, the knowledge base is updated and a new feedback text (reflecting the updated content) is generated. Additionally, spans of feedback text representing an object in the KB can be selected using the mouse to move or remove the object to or from a location in the KB. After each action, a new feedback text is generated representing the updated KB.

Let us consider a simple example that conveys the essential features of WYSIWYM editing. We have a KB consisting of two parts: (1) an ontology in which we specify the set of available concepts and their attributes, and (2) an assertion box (A-box) in which instances of concepts/classes are introduced. Our sample ontology is represented by means of the semantic web compatible OIL language of which we only use a subset:[6]

```
class-def top                 |     class-def account
                              |         subclass-of top
class-def event               |         slotconstraint owner
   subclass-of top            |         value-type person
                              |
class-def person              |     class-def view
   subclass-of top            |         subclass-of event
                              |         slotconstraint agent
class-def client              |         value-type person
   subclass-of person         |         slotconstraint object
                              |         value-type account
class-def employee            |
   subclass-of person         |     class-def transaction
                              |         subclass-of event
                              |         ...
```

We start by introducing the class `top`. We also introduce three subclasses – `event`, `person` and `account` – of `top` and two subclasses of the concept `person`. We have the attribute `owner` for the class `account` and stipulate that its value is a `person`. We `view` is introduced as a subclass of `event`. It has two attributes: `agent` with a `person` as its value and `object` with an `account` as its value.

The A-box contains the actual knowledge to be edited. It can be represented by means of a graph: nodes stand for instances of concepts, i.e., objects, and directed arcs

[6] http://www.ontoknowledge.org/oil

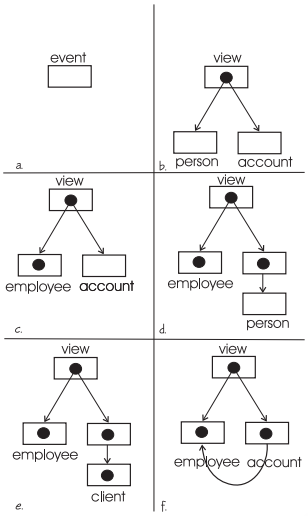

Fig. 2. Editing a Direct Acyclic Graph.

represent attributes. The basic editing operation is that of adding a new object, of a specified type, as the value of an attribute of an existing object.

Let us start with an A-box which expects an instance of the concept `event`; see Figure 2.a. On the basis of this KB a feedback text is generated:

(1) **Something happened**.

The entire span of text is in boldface to indicate that the text is an anchor. By clicking on it, the user obtains a menu showing alternative expansions the KB. Our ontology licenses two options: 1. Some person viewed some account and 2. Some person made a transaction to some account. When the user selects option 1., a new instance of the concept `view` is introduced into the KB (see Figure 2.b). From the updated KB a fresh feedback text is generated:

(2) **Some person** viewed **some account**.

When the user selects the first anchor in this text, the following two options for expanding the KB appear: 1. An employee viewed some account and 2. A client viewed some account.

Our user selects option 1. leading to the new KB in Figure 2.c and the text:

(3) An employee viewed **some account**.

Expansion of the second anchor along the same lines gives rise to the KB in Figure 2.d and the following feedback:

(4) An employee viewed **someone's** account.

If the user expands '**someone's**', the complete network in 2.e can be obtained and the text:

(5) An employee viewed the account of a client.

Instead of inserting a new object ('a client') into the incomplete network (Figure 2.d), the user could have chosen to copy and paste an existing object. The span 'an employee' has a menu with the options `cut` and `copy`. `copy` causes the underlying object to be stored in a buffer. Subsequently, the user can paste it into the incomplete part of the KB, i.e., '**someone's**'. This would result in the network in Figure 2.f. and the following feedback text:

(6) An employee viewed his/her own account.

A reflexive pronoun is generated for the `own` attribute and its a value. Note that if `copy` and `paste` had simply operated on the graphemic level of the sentence instead of the underlying semantics, the result would have been 'An employee viewed an employee account'. The proposed approach is different from, for instance, NLMenu (Tennant et al., 1983) which allows for the menu-based editing of the *syntactic surface structure* of sentences, rather than the *underlying content*.

Coreference is an aspect of meaning which is quite hard to determine automatically but pervasive in dialogue. WYSIWYM avoids this problem by letting the user explicitly specify it during editing. The system avoids interpretation, and thereby also avoids incorrect interpretations. Currently implemented WYSIWYM systems support coreference, and also introduction of plural objects, quantification, part-whole relationships and logical relations such as negation and implication, and tense.[7]

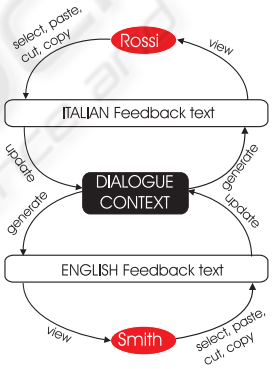

Fig. 3. Crosslingual Joint Knowledge Editing.

[7] http://www.itri.bton.ac.uk/projects/WYSIWYM/wysiwym.html.

5.2 Multi-person Editing and Dialogue

Let us now take the step from single-person editing to multi-person editing. Multi-person editing leads us to crosslingual dialogue. The basic idea is visualized in Figure 3. We have added a second editor with access to the underlying context/KB. Although each editor has access to the same context, their views of it are different: Rossi looks at it through 'Italian glasses' (a language generator for Italian) and Smith through English ones. Of course such a set-up does not necessarily lead to interactions that qualify as dialogues. To approximate dialogue behaviour we introduce some constraints:

1. The jointly edited structure has to be interpreted as representing the *dialogue context* of the dialogue at hand. It consists of the *dialogue history*, progressively built up, and relevant *background information*. This information can be referred to in the course the dialogue and comprises structured objects (e.g., a record with information on a specific transaction, e.g., its date, clients, etc.) and links to information on an intranet or the Internet.
2. Only the most recent turn in the history can be modified, although material can be *copied from preceding turns to establish anaphoric links*.
3. Interlocutors construct turns one at a time.

Figure 4 depicts a snapshot of a conversation between the employees of an Italian and an English bank who use the CROCODIAL technology. Each interlocutor is presented with a WYSIWYM feedback text of the dialogue context at each stage of the dialogue. A common Internet browser is used. In the browser we have a lightweight applet for displaying the mouse-sensitive text with its associated editing operations. The underlying representation and the language generation software for presenting it to the users reside on a central server. In Figure 4, we have italicised some of the phrases whose semantics consists of coreference links in order to illustrate their pervasiveness both inside and across dialogue turns.

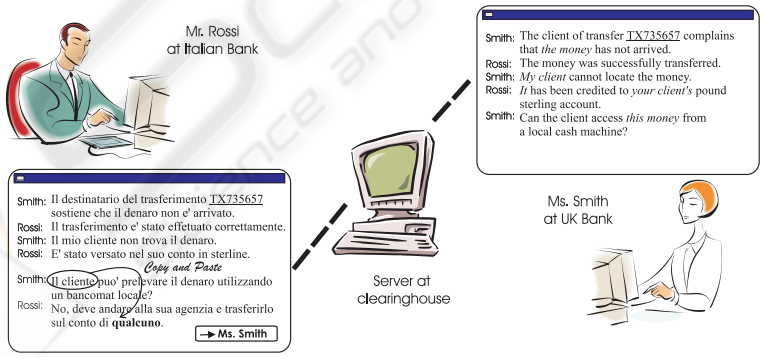

Fig. 4. Crosslingual Joint Knowledge Editing.

In addition to the accuracy and coverage of complexity supported by our approach, it also allows us to benefit from the fact that the interlocutors construct a formal repre-

sentation of the content of the interaction. We propose to exploit this representation by using it to automatically generate a summary. For example, the interaction in Figure 4 could lead a summarizer to produce the following summary which integrates contextual information regarding the transaction (date, banks involved, etc.).

> On 15-1-2003 Ms Smith (Citibank) called Mr Rossi (Banca di Roma) about the transfer of 100,000 GBP to the account of Count Roberto da Silva (654012). It was established that the money had been transferred to the pound sterling account of Da Silva. This account can only be accessed via a local branch of the Banca di Roma.

Similar summaries could be generated on demand in other languages when the need for this arises; the basis for such summaries is the formal representation of the dialogue which the interlocutors unwittingly constructed.

Finally, note that the approach leaves scope for different modes of interaction. A dialogue can be conducted very much like an email interaction with long breaks between contributions, but it can also be conducted in a more synchronous fashion similar to what can be found in a chatroom.

6 Summary and Discussion

The following three features of the proposed system make it suitable for certain practical applications: (1) People with no common language can communicate; (2) Each message is precise and linguistically correct; (3) The content of the conversation is formalized in a knowledge base, so potentially it can be utilized by other programs.

However, the benefits have to be traded off against some limitations: (1) The interaction is text-based, not speech-based; (2) Communication may be slow (compared with face-to-face human-human conversation) because of the time needed to compose contributions by WYSIWYM.

These two limitations would be serious if there was a speech-based alternative that allowed for fast and extremely accurate crosslingual interactions. This is, however, not the case. Firstly, speech-based MT systems still have some way to go before they will attain extremely high levels of accuracy. Secondly, the authors of one of the few systems which does aim for this goal admit that '[...] communication through a translation device is not fast. [...] It is possible for the component technologies (recognition, translation and synthesis) to become more streamlined, but it would be very difficult to achieve truly spontaneous, simultaneous translation.' (Frederking et al., 2002)

In fact, we feel that it is misguided to present current speech-based MT as a competitor of the CROCODIAL approach. Firstly, there are many applications in which extreme accuracy is not called for. Secondly, we see potential for *hybrid* solutions. Some translation systems provide a so-called back translation. If such a back translation were based on an interlingua, it would be possible to use our approach to correct the back translation whenever necessary, by means of WYSIWYM editing. This could allow interlocutors to circumvent cumbersome clarification dialogues.

The enabling WYSIWYM technology has been applied to a number of domains. A version of the system is available to the research community (Evans & Power, 2003). We are currently building a first CROCODIAL prototype for a small financial domain. A

number of preliminary evaluations of WYSIWYM have been carried out. These studies have indicated that users find the WYSIWYM editing operations and feedback to lead to predictable results and follow a logical pattern. However, it has also been established that an incomplete ontology negatively affects user satisfaction. Currently, more elaborate evaluations with eye-tracking equipment are in progress at the Evaluation Centre of the German Research Centre for Artificial Intelligence.[8]

The approach we have described is grounded in natural language technology; in order for it to work, we need a generator that maps the formal representation of the context to a natural language text (existing WYSIWYM generators cover English, German, French and Italian). Each generator for a new language extends the scope of the technology. Unfortunately, existing language generators are not readily reusable because they require widely varying inputs. However, the emergence of the semantic web is likely have a positive impact: many systems already have the ability to use XML input, and content representation languages, such as OIL, may turn out to be a first stepping-stone towards standardization.

References

1. Evans, R., Power, R.D. (2003). WYSIWYM - Building Interfaces with Natural Language Feedback, Proceedings of EACL03 (Demonstrators), Budapest, 203–206.
2. Frederking, R., Black, A., Brown, R., Moody, J. and Steinbrecher, E. (2002). Field Testing the Tongues Speech-to-Speech Machine Translation System. *LREC2002*, Las Palmas, Canary Islands.
3. Hutchins, J. (1999). Retrospect and prospect in computer-based translation. In *Proceedings of* MT *Summit VII*, 13th-17th September 1999, Kent Ridge Digital Labs, Singapore, 30–34.
4. Isard, S. (1975). Changing the Context, In: E. Keenan (Ed.), *Formal Semantics of Natural Language*, Cambridge: Cambridge University Press, 287–296.
5. Lonsdale, D.W., Franz, A.M., Leavitt, J.R.R. (1994). Large-scale Machine Translation: An Interlingua Approach. In *IEA/AIE '94: Proceedings of the Seventh International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, May 31-June 3, 1994, Austin, Texas. ACM, 1994.
6. Power, R., D. Scott and R. Evans (1998). What You See Is What You Meant: direct knowledge editing with natural language feedback, *Proceedings of ECAI-98*, Brighton, UK, 1998, 180–197.
7. Rayner, M., D. Carter. P. Bouillon, V. Digalakis & M. Wirén (2000). *The Spoken Language Translator*, Cambridge, Cambridge University Press.
8. Tennant, H., Ross, K., Saenz, M., Thompson, C., & Miller, J. (1983). Menu-Based Natural Language Understanding. In *Proceedings of the 21st Annual Meeting of the Association for Computional Linguistics*, Cambridge, Massachusetts, pages 151–158.

[8] http://www.dfki.de/LT-EVAL/Seiten/Englisch/Work/wysiwym.htm