A system for generating teaching initiatives in a computer-aided language learning dialogue

Nanda Slabbers
Dept of Computer Science
University of Twente

Alistair Knott
Dept of Computer Science
University of Otago

Abstract

This paper describes an extension made to a bilingual human-machine dialogue system, to allow the system to take initiatives in a language-learning dialogue. When the user concedes the initiative to the system, the system generates a set of 'possible initiatives', and chooses the best of these based on a number of criteria. These criteria relate firstly to the *formal* goal of generating an initiative which is appropriate in the current context, and secondly to the *substantive* goal of teaching the student a set of targeted syntactic constructions.

1 Introduction

A system engaging in a dialogue with a user has to generate two quite different kinds of utterances: **responses** (such as acknowledgements, answers to questions, and clarification questions) and **initiatives** (such as assertions of new material, or new questions à propos of nothing). When we consider what is involved in these two kinds of utterance, there are some interesting differences. It is common to analyse the task of natural language generation (NLG) as a pipeline involving **content selection**, **sentence planning** and **syntactic realisation** (see e.g. Reiter, 1994). For the generation of responses, the task of content selection is normally simple; the burden of the work is in sentence planning and syntactic realisation. For instance, to generate answers to questions or clarification questions, we typically need to construct sentences whose syntax and semantics echo that of the sentence being responded to. For the generation of initiatives, on the other hand, content selection is a key process: the issue of 'what to say' is much less constrained for such utterances. In this paper, we describe a system for generating initiatives in a particular register of dialogue: **computer-aided language learning** (or **CALL**) dialogue. The main innovation in our system is its adaptation of some standard content-selection techniques from NLG (traditionally used to produce utterances in monologue) to the task of generating initiatives in such dialogues.

We will begin in Section 2 by surveying some existing systems which generate teaching initiatives. Section 3 describes the initiative module and its goals. Section 4 describes the dialogue system in which the initiative module is embedded and provides some

results.

2 Existing work in generating teaching initiatives

There has been a great deal written about the role of initiative in tutorial dialogue systems; see e.g. Haller and McRoy (1997). But comparatively little of this work has considered the situation where the topic being taught is a foreign language. A CALL dialogue need not resemble a tutorial interaction at all; in many cases, it simply looks like a (somewhat stilted) conversation between two speakers on a particular topic. Of course, either participant can also ask or answer explicit questions about the language being taught. But when the topic being taught is the language itself, simply advancing the conversation has educational merit in its own right. The initiatives made by the tutor thus have a dual function: to continue a natural-sounding conversation, and to do so in a way which scaffolds the student's current language learning.

Surprisingly, most dialogue systems specialising in language-learning do not focus on generating initiatives. The systems we have reviewed (e.g. Desmedt, 1995; Seneff *et al.*, 2004; Raux and Eskenazi, 2004) typically involve a scenario where the user has to accomplish some task, and in which therefore most initiatives come from the student. In these scenarios, it is hard for the student to learn by adapting utterances made by the teacher. In our system, we focus on more symmetrical dialogues where the student and tutor can make the same kinds of utterances (e.g. 'Where's your Mum from?'... 'Where's *your* Mum from?'). In these dialogues, the tutor's initiatives can provide the student with models of the constructions to be learned, as well as fleshing out the content of the dialogue. The question is: how to generate appropriate initiatives in such contexts? We believe that some standard content selection techniques from NLG can be usefully applied to the problem.

3 NLG content-selection methods for initiative generation

The process of content selection in NLG is typically defined in relation to two goals: firstly the **formal** goal of generating a coherent text, and secondly the **substantive** goal of achieving a certain effect on the hearer. If the text being generated is a monologue, the formal goal will be expressed in terms of a theory of discourse structure, such as RST or one of its many competitors. The substantive goal is typically expressed using the vocabulary of AI planning. In one common architecture for content selection (see e.g. Marcu, 1996; O'Donnell *et al.*, 2001), the process involves two passes. In the first pass, a large set of **candidate messages** is created, using heuristics designed to maximise the likelihood of achieving the system's formal and substantive goals. In the second pass, these candidate messages are evaluated more systematically, and the one which best achieves the goals is chosen to be generated. This more systematic process often involves 'look-ahead' to the sentence planning and realisation stages, so that the evaluation can take into account syntactic factors as well as semantic ones.

To adapt the model just outlined to the generation of initiatives in a CALL dialogue, we must first specify formal and substantive goals for the system, and then we must specify a procedure for generating and evaluating initiatives in relation to these goals. These topics will be considered in the remainder of this section.

3.1 Formal goal: dialogue coherence

The formal goal of our CALL system will be to maintain a coherent dialogue. Modelling di-

alogue coherence is hard; it is not possible to define coherence at the level of dialogue acts (e.g. 'a question begets an answer'), because in the general case, the semantic content of a dialogue act is as relevant as its type. Current models of dialogue coherence typically use some brand of update semantics to formalise different dialogue acts and to provide definitions of grounding, the relationship between questions and answers, and subdialogues (c.f. e.g. Traum *et al.*, 1999). However, while these complexities are necessary in order to constrain *response* dialogue acts, they do not seem so necessary for initiatives. If we restrict ourselves to contexts where an initiative must be taken, it seems possible to define a coherent dialogue move simply by enumerating the types of dialogue act which can be taken at this point. In our case, we introduce two special dialogue acts which can only be used to make initiatives: a **new assertion** (which we distinguish from assertions which provide the answers to questions) and a **new question** (which we distinguish from clarification questions and follow-up questions). For our CALL domain, we decompose new questions into **genuine questions** (which fill in gaps in the system's knowledge base) and **teaching questions** (which ask the student about information already in the common ground, to check whether it has been understood).

In addition to this restriction to particular dialogue act types, we posit two weaker formal criteria for initiatives. The first relates to the **topic** of the new utterance. We suggest there is a preference for initiatives which maintain the current topic of the dialogue. At some points topic *changes* may be preferable instead (especially when the dialogue is on the same topic for a long time), but we assume the student will change the topic when he wants to. We do not see topic continuity as essential for maintaining coherence, but certainly if there is no continuity, there are obligations to mark this textually in the utterance generated. In our model, the topic of an utterance is the set of individuals and predicates which it introduces, and the degree of topic continuity between two utterances is defined in terms of the overlap between the two relevant sets; see Slabbers (2005) for details. There is also a higher-order preference for **strategic** initiatives, which move onto a topic which the system knows a lot about. The system is configured to prefer assertions which introduce topics which appear frequently in the its private knowledge base of facts. The second weaker criterion relates to the **mix** of dialogue acts; we suggest there is a preference for interleaving dialogue acts of different types, rather than producing several acts of the same type. Dialogue act mix is a **global** constraint on dialogue coherence (in the sense of Hovy, 1988; Piwek and van Deemter, 2003) but nonetheless it is one which we can try and optimise locally. In cases where several candidate utterances score equally as regards topic continuity, we can give preference to those which realise dialogue acts which have not been recently used.

3.2 Substantive goal: language-learning

In a CALL dialogue, any initiative made by the system should further its goal of teaching the student the language. Since our dialogue system creates complete syntactic representations both when parsing student input and when generating teacher output, we can specify the system's educational goal very precisely, as a set of **target syntactic rules**. We assume that the system will deliver a sequence of dialogue-based lessons, beginning with dialogues featuring simple syntactic constructions and progressing in each subsequent dialogue to more complex constructions. The substantive goal of each lesson is for the student to show evi-

dence of understanding the rules 'featured' in the lesson; utterances which involve featured rules (or which are likely to elicit them) can then be scored higher than those which do not.

3.3 The initiative generation algorithm

Our algorithm for generating initiatives has four steps. First, we identify a set of possible topics for the new initiative. During the second step of the algorithm we generate a set of **candidate messages** of each dialogue act type: new assertions, genuine questions and teaching questions. A separate algorithm is used in each case, comprising content selection and sentence planning phases, but stopping short of syntactic realisation. (The algorithms for generating new assertions and genuine questions require the system to have a private knowledge base of facts and question-generation rules; see Section 4.1 for how this is created.) The algorithm for generating teaching questions selects a fact from the common ground and turns it into a yes-no question or a wh-question by manipulating its logical form. The result of these algorithms is a set of candidate messages, each represented as a logical form. We then consult a history of previous system utterances, and discard any initiatives which have previously been generated by the system, whether as initiatives or responses, so that the system never repeats itself when taking an initiative.

The third step of the algorithm consists of scoring the remaining initiatives on a range of different criteria: all initiatives get scores for the suitability of the dialogue act (based on the mix of the previous dialogue acts), the degree of topic maintenance, and finally a dialogue-act-specific score determined in different ways for each different dialogue act. Assertions get a score based on the strategy criterion (e.g. initiatives about topics which the system knows a lot about receive a higher score); teaching questions get a score based on the complexity of the question, with more complex questions being preferred; and genuine questions get a score based on the order in which question-formation rules were entered by the author (see Section 4.1), which reflects the author's view of their importance. The scores are normalised, summed and ranked to create a **shortlist** of initiatives. Finally, each initiative on the shortlist is passed to the sentence generator, and a second evaluation is carried out which assesses to what degree sentences use syntactic rules which have not yet been assimilated by the student. The winning initiative is delivered to the user.

4 Initiative generation in the Kaitito dialogue system

Our dialogue system, called Te Kaitito[1], supports bilingual written human-machine dialogues in English and Māori, the indigenous language of New Zealand. The Te Kaitito CALL system is originally meant to teach Māori, but it has a modular design, and can work to teach any language for which a grammar is specified. In this paper we will use the English grammar, so the system should be viewed as a CALL system for English.

The user and the system alternate in generating contributions to a dialogue. When it is the user's turn to contribute, she enters a sentence in English. The sentence is first parsed, using the LKB system (Copestake *et al.*, 2000) and the ERG grammar (Copestake and Flickinger, 2000). The parser produces a set of syntactic analyses, each of which can have several semantic interpretations after its presuppositions have been resolved against the common ground. One interpretation is then selected, using a combination of disambiguation tech-

[1] Online demos of Te Kaitito can be found at `http://tutoko.otago.ac.nz:8080/teKaitito/`.

niques (see Lurcock *et al*, 2004). The dialogue manager then determines how to create a message in reply—either using a 'response' dialogue act, or by invoking the initiative module. In either case, the response message is passed to a sentence planner for computing referring expressions and discourse signals, and then to a sentence generator. The generator consults the same grammar used by the parser to create the text which is returned to the user.

4.1 Authoring mode dialogues

In order to be able to generate initiatives, a dialogue system needs to be given a knowledge base of private information, on which to draw to create assertions and questions, and a set of substantive goals in relation to which candidate initiatives can be evaluated. In our system, both the knowledge base and the substantive goals are created during a special kind of dialogue with the system called an **authoring dialogue**. In authoring mode, the user is assumed to be a teacher, creating a lesson plan for the system. An example dialogue is given in Figure 1. The system begins with an empty common ground. The teacher authors a character by telling the system facts about itself (e.g. Utterance 1), and by entering **question-generation rules** specifying what kinds of question to ask about different types of objects (e.g. Utterance 2). (The assumption is that the author will enter rules in order of decreasing priority. When ranking alternative candidate initiatives, therefore, the system will prefer a question derived from application of a rule authored earlier during authoring mode.) At the end of an authoring dialogue, the system saves the set of facts in its common ground into a private knowledge base, and saves the set of question-generation rules into a separate private knowledge base. It also automatically creates a set of target syntactic rules for the lesson (see Section 3.2), by traversing the parse trees for every utterance in the authoring dialogue and recording all the rules which are used in this dialogue but not in the authoring dialogues for previous lessons.

4.2 Student mode dialogues

The start of the dialogue

When the system enters **student mode**, its common ground is initialised to empty, and it loads a private knowledge base and agenda of rules created by one of the authoring dialogues. It then enters a conversation with the student.

During the dialogue

Once a dialogue has been initiated, the system and the student alternate in making contributions to the dialogue. The dialogue consists of pairs of forward-looking and backward-looking dialogue acts—for instance, assertions and (possibly implicit) acknowledgements, or questions and answers. At the end of any such pair is a **transition relevance point—**a point where either participant can take an initiative. At such points, the system always passes the initiative to the student. However, the student can concede the initiative, simply by hitting <return>.

An example of a mixed-initiative dialogue based on the facts and goals created in the authoring dialogue in Figure 1 is given in Figure 2. The system begins by generating all possible initiative messages in the null context, and computing their normalised scores. In the initial context, with an empty common ground, genuine questions are preferred over assertions, and teaching questions are impossible. The candidate initiatives whose score exceeds a given threshold are passed to the sentence generation module, and they receive a syntactic score, which is combined with their original scores to generate a total score, as shown in Figure 3. The possible initiatives are

Auth	Your name is John. (1)	Auth	Ask: how is a person? (2)
TK	Okay.	TK	Okay.
Auth	You are from Dunedin.	Auth	Ask: what is a person's name?
TK	Okay.	TK	Okay.
Auth	You are fine.	Auth	Ask: where is a person from?
TK	Okay.	TK	Okay.

Figure 1: A simple authoring dialogue

TK	How are you? (1)	TK	I am John.
S	I am fine.	S	<return>
TK	Okay.	TK	Where are you from? (6)
S	How are you? (2)	S	I am from Dunedin.
TK	I am fine.	S	<return> (7)
S	<return> (3)	TK	I am also from Dunedin. (8)
TK	What is your name? (4)	S	<return>
S	My name is Ann.	TK	Let's see if you remember: Who am I? (9)
TK	Okay.	S	You are John.
S	Who are you? (5)	TK	Correct, well done! (10)

Figure 2: A simple mixed-initiative dialogue

```
DIAL-ACT SCORE TOPIC SCORE QN-ORDER SCORE  SYN. SCORE  TOTAL  SENTENCE
100            100         100              78         378    how are you
100            100          67             100         367    what is your name
100            100         100               3         303    you are how
100            100          67               3         270    your name is what
100            100          33              78         311    where are you from
100            100          33              56         289    from where are you
```

Figure 3: Normalised scores of initiatives in the shortlist for Utterance 1

all genuine questions, so the initiatives receive scores for the following criteria: a dialogue-act score based on the mix of the different dialogue acts, a topic score based on the similarity with the current topics, a question order score based on the order in which the genuine questions were entered by the author, and a syntax score based on the syntactic rules used in the initiative. On this basis, the first initiative generated (Turn 1) is *How are you?*. (Note that an alternative realisation of this sentence, *You are how?* scores badly at a syntactic level, because it involves several rules not used in the authoring dialogue.) If the student does not answer this question as expected the initiative will be repeated. However, in this example the student does answer the question as expected, so the dialogue continues normally. Next, the student is offered the initiative again, and she decides to ask the system a similar question (Turn 2), which the system answers. Then the student concedes the initiative (Turn 3), and the system asks the next-best genuine question (Turn 4). The student answers this, and then asks a similar question in response (Turn 5). The system then asks its last genuine question (Turn 6). When the student responds, and then concedes the initiative again (Turn 7), the system generates a new assertion on the current topic (Turn 8). Finally, when the student again concedes the initiative, the system opts to generate a teaching question (Turn 9), and when the student answers correctly, it provides some positive feedback (Turn 10).

The end of the lesson

The dialogue continues until the system has evidence that each of the target constructions has been assimilated by the user. This evidence comes in a number of forms; for instance, if the user correctly answers a teaching question, the system increments the assimilation score for each rule in both the question and its answer. When all rules have been assimilated, the lesson ends successfully, and the student is allowed to proceed to the next lesson. Sometimes it may happen that a student does not learn a new rule even when (s)he is shown an instance of it being correctly applied. Given that the system never repeats itself when taking an initiative, it might therefore happen that there are no candidate initiatives which will help assimilate any of the remaining unassimilated target rules. In such a situation the lesson ends unsuccessfully, and the student is asked to consult the teacher.

5 Conclusions and future work

Informal evaluations suggest that Te Kaitito's teaching dialogues provide a useful environment in which a student can practice conversational skills in the language being learned. The system's grammar, and repertoire of dialogue moves, are naturally very simple. But in a language-learning environment, particularly for novice language learners, this limited coverage is not as harmful as it normally is. The student's own grammar and vocabulary are similarly limited, and if we know which textbook is being used, and what stage in the book (s)he is at, we have a good chance of being able to build a grammar which can handle all the constructions (s)he is likely to attempt.

We believe that addding initiatives to the language-learning dialogue is very beneficial. If the student is lost, hitting <return> is a simple way of progressing the dialogue (though naturally there are still some places where the student has to respond with something other than <return>). And the initiatives taken by the system create models of well-formed sentences which the student can modify and try out him/herself. In a forthcoming evaluation, we will test more formally whether this is the case.

Naturally there are many aspects of CALL dialogues which we are not yet simulating in the current work. Most obviously, while we are generating teaching *initiatives*, we do not yet generate teaching *responses*—i.e. utterances whose aim is to alert the student to a mistake that (s)he has made, and to provide assistance in correcting the mistake. This is something we are considering in current work.

Acknowledgements

This research was funded by the NZ Foundation for Research, Science and Technology (FRST) grant UOOX0209.

References

Copestake, A. (2000). The (new) LKB system. CSLI, Stanford University.

Copestake, A. and Flickinger, D. (2000). An open-source grammar development environment and broad-coverage English grammar using hpsg. In *Proceedings of LREC 2000*, Athens, Greece.

Desmedt, W. (1995). Herr Kommissar: An ICALL conversation simulator for intermediate German. In M. Holland, J. Kaplan, and M. Sams, editors, *Intelligent language tutors: Theories shaping technology*. Lawrence Erlbaum Associates, Mahwah, NJ.

Haller, S. and McRoy, S. (1997). Papers from the AAAI Spring Symposium on Computational Models for Mixed Initiative Interactions. AAAI Technical Report SS-97-04.

Hovy, E. (1988). *Generating natural language under pragmatic constraints*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Lurcock, P., Vlugter, P., and Knott, A. (2004). A framework for utterance disambiguation in dialogue. In *Proceedings of the 2004 Australasian Language Technology Workshop (ALTW)*, pages 101–108, Macquarie University.

Marcu, D. (1996). Building up rhetorical structure trees. In *Proceedings of the AAAI annual meeting*.

O'Donnell, M., Mellish, C., Oberlander, J., and Knott, A. (2001). ILEX: an architecture for a dynamic hypertext generation system. *Natural Language Engineering*, **7**.

Piwek, P. and van Deemter, K. (2003). Dialogue as discourse: Controlling global properties of scripted dialogue. In *AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, Stanford, CA.

Raux, A. and Eskenazi, M. (2004). Using task-oriented spoken dialogue systems for language learning: Potential, practical applications and challenges. In *Proceedings of the InSTIL/ICALL Symposium*, pages 147–150.

Reiter, E. (1994). Has a consensus nl generation architecture appeared, and is it psycholinguistically plausible? In *Proceedings of the 7th. International Workshop on Natural Language generation (INLGW '94)*.

Seneff, S., Wang, C., and Zhang, J. (2004). Spoken conversational interaction for language learning. In *Proceedings of the InSTIL/ICALL Symposium*, pages 151–154.

Slabbers, N. (2005). A system for generating teaching initiatives in a computer-aided language learning dialogue. Technical Report OUCS-2005-02, Department of Computer Science, University of Otago, Dunedin, New Zealand.

Traum, D., J, B., Cooper, R., Larsson, S., Lewin, I., Matheson, C., and Poesio, M. (1999). A model of dialogue moves and information state revision. TRINDI project deliverable.