

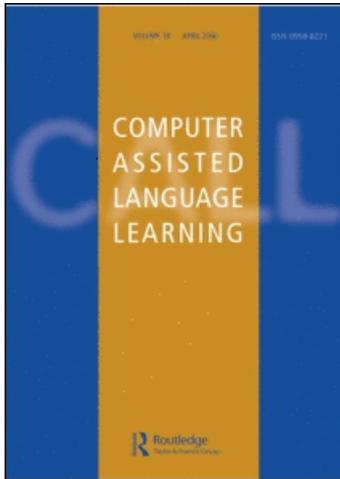
This article was downloaded by: [Radboud Universiteit Nijmegen]

On: 12 May 2009

Access details: Access Details: [subscription number 908165048]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Computer Assisted Language Learning

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t716100697>

Let's Chat: A conversational dialogue system for second language practice

Iain A. D. Stewart^a; Portia File^a

^a University of Abertay Dundee, UK

Online Publication Date: 01 April 2007

To cite this Article Stewart, Iain A. D. and File, Portia(2007)'Let's Chat: A conversational dialogue system for second language practice', *Computer Assisted Language Learning*, 20:2, 97 — 116

To link to this Article: DOI: 10.1080/09588220701331386

URL: <http://dx.doi.org/10.1080/09588220701331386>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Let's Chat: A conversational dialogue system for second language practice

Iain A. D. Stewart* and Portia File

University of Abertay Dundee, UK

Early and intermediate second language (L2) learners often encounter difficulties when engaging in introductory social conversations, typically having had little opportunity to practise such interactions. This article describes a project to design and prototype a computer dialogue system, Let's Chat, which would allow learners to rehearse social conversations without a human partner. While most existing systems do not facilitate interactions of a sufficiently high quality to be useful for L2 acquisition, Let's Chat does so by relying on communication through pre-stored phrases, thereby steering clear of the technological limitations of conventional natural language processing. This approach is supported by research that suggests that the holistic assimilation of L2 sequences can boost the communicative capabilities and fluency of learners. By practising with Let's Chat, learners would acquire good target language input in the form of well-constructed, native-like utterances and thus be better-prepared, more confident participants during natural conversations.

Introduction

Many, perhaps most, early and intermediate second language (L2) learners struggle when attempting to engage in introductory social conversations in the target language. The almost universal acceptance of communicative approaches to language teaching has undoubtedly resulted in enhanced levels of conversational proficiency among learners, but much classroom practice focuses on transactional scenarios, such as ordering a meal in a restaurant, making an airline reservation, or speaking to a shop assistant. Rehearsing everyday social chats is a low

*Corresponding author. 13 Harestone Crescent, Downfield, Dundee, DD3 0LQ, Scotland, UK.
Email: iain.a.d.stewart@btinternet.com

priority in many L2 curricula and learners are often left to manage this for themselves.

Given that dialogue is recognized as a “useful environment for language learning” in which “the medium of the interaction is itself the topic being learned” (Vlugter, Knott, & Weatherall, 2004, p. 1), perhaps the computer could gainfully offer a relaxed, entertaining environment for practise of the social conversation skills regularly neglected in classroom drills. However, very few computer-assisted language learning (CALL) dialogue systems have addressed the issue of social conversation in any major way, typically focusing instead on more restricted, and hence easier to handle technically, domains involving interactions of a context-specific, chiefly transactional nature (see, for example, Lewis Johnson *et al.*, 2004).

Technological Limitations

While quite substantial progress has been made in recent years in the development of effective speech recognition and synthesis software, the same cannot be said of natural language processing. Indeed, as Wilks and Catizone note, “practical and theoretical investigations into the nature and structure of human dialogue have been a topic of research in artificial intelligence and the more human areas of linguistics for decades: there has been much interesting work but no definitive or uncontroversial findings” (2001, p. 1). They go on to assert that “the best performance overall in HMC (Human–machine conversation) has almost certainly been Colby’s PARRY program since its release on the (then ARPA) net around 1973” (p. 1). Even the latest dialogue systems, such as the “chatterbot” ALICE (<http://www.alicebot.org/>), multiple winner of the Loebner Prize (<http://www.loebner.net/Prizef/loebner-prize.html>) for the most human-like computer and with which the user communicates exclusively by text, fall well short of passing the benchmark Turing test, the Holy Grail of artificial intelligence (AI) that requires a system to fool an observer into believing that it is itself human.

In terms of L2 practice, the ideal scenario would be for a learner to speak to the computer and for the computer to “understand” and respond in a sufficiently appropriate and native-like manner to provide good target language input. But is this a realistic aim given the current state of natural language processing? As Feigenbaum has observed, real difficulty lies in managing “the ‘understand’ part: the semantics that attach real-world meaning to the word-symbols, then use those meanings for knowledge organization and inference” (2003, p. 33). Another eminent researcher in the field of AI believes that the dream of a system that is truly able to understand may never be realized (Charniak, 1995).

An Alternative Approach to Human–Computer Dialogue

If existing dialogue systems cannot enable a sufficiently naturalistic and flexible interaction to be of value to L2 learners, and given that there is little reason to predict

an imminent improvement in their capabilities, is there an alternative to traditional natural language processing that could be employed to better effect? We believe that facilitating dialogue through pre-stored utterances is an option that not only overcomes current limitations in the artificial understanding of natural language but may also be more appropriate for a system designed to function as an aid to second language acquisition (SLA).

Whereas a typical chatterbot enables the user to communicate by freely typing his/her input, a dialogue system that relies on pre-stored phrases requires the user to select from a range of utterances displayed on the screen. Within such a system, there is no need for conventional natural language processing (cleaning of input, parsing, etc.), as each phrase is stored in the underlying system database alongside descriptive data, such as utterance type (e.g. question, reply, story), topic, subtopic and its possible roles within defined conversation models. When the user selects an utterance from the interface, a query is generated by the "dialogue engine" and sent to the database to seek a suitable system response that is defined by the attributes of the user input. By eliminating the need for detailed linguistic analysis of the user input and the associated difficulties of constructing an adequate reply, such a system allows for high-quality interactions that escape the stilted, contrived feeling of most chatterbot conversations.

The fact that the user's choice of utterance is restricted to those already contained within the database can be seen as a limitation of a pre-stored phrase dialogue system. Rather than perceiving this as a drawback, we would argue that communication through pre-stored material is potentially advantageous in an L2 context, a view founded upon research discussed below into the nature of formulaic language and its role in the SLA process.

Pre-stored Utterances and Second Language Acquisition

Formulaic Language

Pre-stored utterances can be described as essentially formulaic in nature. While the exact determinants of formulaic language are open to dispute, a common starting point is to equate formulaicity with an absence of creativity or novelty in the construction of a phrase (Wray, 2000, p. 45). When such phrases are resistant to lexical analysis on account of possessing a meaning unrelated to the individual words of which they are composed ("to buy the farm" or "to kick the bucket" as euphemisms for "to die", for example), it is uncontroversial to classify them as formulaic. Other word strings may appear at first glance to be perfectly regular and semantically transparent yet still be deemed formulaic by some observers through possessing a holistic nuance beyond their component parts, especially to members of the speech community that engendered them (Wray, 2002). In more general terms, however, any word sequence that the speaker retrieves whole from memory, rather than constructs from its constituent parts on an *ad hoc* basis, may be regarded as formulaic.

However one characterizes formulaicity, it is clear that a substantial part of everyday speech falls within the boundaries of its broadest definition; as Fillmore noted in 1976, “an enormously large amount of natural language is formulaic, automatic, and rehearsed, rather than prepositional, creative, or freely generated” (cited by Vihman, 1982, p. 261). A number of compelling explanations for the prevalence of formulaic structures are advanced by Wray (2000, 2002, especially chapter 5), which, in broad terms, can be perceived as fulfilling the following two functions: (1) that of aiding the speaker’s production by lightening the processing burden and thus facilitating fluency; and (2) that of aiding the hearer’s comprehension of the full message the speaker wishes to convey, including implicit assertions such as those of individual and/or group identity. Thus, Wray contends, all formulaic sequences ultimately serve to promote the speaker’s interests (2002, p. 95).

Formulaic Language in SLA

While the importance of formulaic language in both the first and second language acquisition processes has been recognized for some time (see, for example, Vihman, 1982 and Weinert, 1995), probably the most thorough and persuasive analyses are those advanced by Wray in recent years (1998, 1999, 2000, 2002). With specific reference to SLA, Wray observes:

In its narrowest sense, formulaicity has always been a useful entrance point for the learner, with the “phrase-book” approach providing a few pre-learned utterances for asking the way to the station or ordering a cup of coffee. However, there is much more to it than this. Gaining full command of a new language requires the learner to become sensitive to the native speakers’ preferences for certain sequences of words over others that might appear just as possible. From the bizarre idiom, through the customary collocation, to the turns of phrase that have no other apparent linguistic merit than that “we just say it that way”, the subtleties of a language may floor even the proficient non-native, not so much because of a non-alignment between interlanguage and target language forms, as because the learner lacks the necessary sensitivity and experience that will lead him or her unerringly away from all the grammatical ways of expressing a particular idea except the most idiomatic. (2000, p. 463)

Drawing on the model of SLA proposed by Wray (2002), itself based on a wide-ranging survey of studies by other researchers, the acquisition process can be divided into three major phases:

1. the early stage, during which formulaic sequences are commonly promoted and are assimilated holistically, without the learner being required to comprehend fully their grammar and syntax;
2. the intermediate and advanced stage, during which learners turn away from formulaicity in the quest to attain a more varied and adaptable communicative capacity; and

3. the stage where the learner is approaching native or near-native competence, at which time holistically acquired utterances regain crucial importance in the pursuit of idiomatic, truly native-like expression.

Learners in the second phase of this model typically rely on creation of novel utterances; while many of these may be grammatically correct, they are effortful to produce, thereby reducing fluency, and are often distinctly non-native-like. A further problem is that expressions acquired holistically in the first phase of the SLA process are commonly distorted when used by second-phase learners. Although in some cases this may simply be the result of defective initial learning, it seems more probable that progression towards analytic capacity leads learners to revisit sequences previously acquired holistically and modify them incorrectly or unidiomatically. This is particularly true of those expressions that have not been rehearsed regularly throughout the learning process; as Wray notes, “holistically learned strings are, by definition, subject to loss of detail, because they rely on the memory of the visual and/or phonological shape of the entire unit” (2002, p. 200).

If we accept that the goal of language teaching is to educate learners to attain a grammatically accurate, fluent and idiomatic command of the target language, then there is a strong case for encouraging the holistic assimilation of formulaic sequences and their frequent rehearsal during all of the three phases of the acquisition process described above. As the holistic approach is already in common use during the earliest phase, the evident challenge to L2 tutors is to find ways of preserving the previously acquired sequences and adding to them in the subsequent stages. The particular requirement is to combine formulaicity with the prevailing emphasis on analyticity during the second phase; neither on its own will lead to mastery of the target language. Excessive focus on formulaicity imparts insufficient grammatical knowledge, yet all-out analyticity does not make for native-like fluency and idiomaticity. As Wray (2000, p. 484) suggests, there is a need for heightened “sensitivity” on the part of the syllabus designer to the intricacies of formulaic language in order to fully exploit its potential benefits in SLA.

Effectiveness of Pre-stored Utterances in L2 Social Conversation

As part of an earlier project, one of the authors conducted an investigation into the quality of L2 social conversations mediated entirely through pre-stored utterances (Stewart, 2004). Four dyads of intermediate learners of Spanish participated in the study. Each dyad carried out six learner-to-learner topic-oriented conversations (introductions, family, friends, hobbies, holidays and studies), conversing naturally for three randomly selected topics and communicating using only phrases pre-stored in a computer communication device, Speak-Spanish, for the remaining three topics. Prior to each conversation, the researcher worked with the participants to develop a range of suitable utterances, which, in the case of computer-mediated topics, were then entered into the SpeakSpanish system.

All conversations were observed by an experienced L2 evaluator, who scored them across three key measures of linguistic competence (command of vocabulary and idiom; clarity of communication; fluency) using a five-point Likert scale (1 = very poor; 5 = excellent). The six conversations of each dyad were then ranked according to their respective aggregate scores and the results analysed using a randomization test, a valid statistical procedure for small-scale experiments (see Todman & Dugard, 2001). In three of the four dyads, all conversations using the SpeakSpanish system were ranked higher than the natural interactions ($p = 0.05$ for each replication). In the remaining dyad, SpeakSpanish conversations were ranked 1 and 2, but with a natural interaction taking third place ($p = 0.45$). When the results of all four dyads were combined, a highly significant result was obtained in favour of computer-aided communication by pre-stored phrases ($p < 0.01$).

The result in favour of computer-mediated communication by pre-stored phrases was underlined by comparing the means of the qualitative scores assigned by the evaluator: conversations mediated through SpeakSpanish received a mean aggregate score of 11.75, whereas the mean aggregate score for those carried out using natural language was an appreciably lower 8.33. Looking at a breakdown of the aggregate scores by evaluation criteria (command of vocabulary and idiom; clarity of communication; fluency), there were very few instances in which natural language conversations scored more highly in any category than those mediated by SpeakSpanish.

Let's Chat: Human – Computer Social Conversation through Pre-stored Utterances

System Goals

So far, we have presented a technical argument in favour of a pre-stored phrase dialogue system, a theoretical justification for the promotion of pre-stored, or formulaic, sequences as part of the SLA process, and experimental data showing the capacity of pre-stored utterances to enable good social conversations between intermediate L2 learners. Since September 2004, the authors have been leading a project at the University of Abertay Dundee, Scotland, with the aim of designing and prototyping a pre-stored phrase dialogue system that would allow L2 learners to prepare for everyday social conversations in an attractive, stress-free environment without the need for a human partner to be present. By practising role-playing interactions with the system across a variety of themes, learners would receive input in the form of correct, idiomatic utterances, with the result that they should become able to draw upon these expressions confidently and fluently during future natural language conversations.

As an initial step, we compiled an outline requirements specification, which serves as a succinct summary of our vision.

1. The system will offer an engaging environment to help learners of a second language develop conversational and dialogue management skills across a range of everyday topics.
2. The initial prototype will focus on English on the grounds that it is the one language shared by all members of the project team, but the system will be designed in such a way as to be easily adaptable to most other commonly studied languages. Aside from minor modifications to the user interface, this will be achieved by “plugging in” a new database of utterances. (It is recognized that the different pragmatic and cultural constraints applying to other languages would have to be respected, but these would affect the creation of conversation material much more than any of the system’s technical capabilities.)
3. Conversations will resemble authentic, interpersonal dialogues as far as possible. However, some restrictions on conversation structure will be necessary to facilitate implementation and to fulfil the system’s tutorial purpose.
4. Users will communicate with the system by selecting from a range of pre-stored utterances displayed on the screen.
5. In addition to utterances specific to particular topics, users will be able to interject generic comments into the conversation at appropriate moments. Users will do this by pressing a button denoting the type of comment to be made and a suitable utterance will be randomly selected from those stored, e.g. clicking on the “Good” button might generate approximately interchangeable phrases such as “That’s good”, “Great”, or “Excellent”.
6. The system will communicate with users by displaying utterances on the screen, which will also be voiced via a text-to-speech engine.
7. When the user selects an utterance, the system will process this input and respond accordingly.
8. If the user selects an appropriate utterance, the system will move the conversation forward.
9. If the user selects an utterance that is contextually inappropriate, the system will detect this, inform the user and offer the opportunity for a new selection to be made.
10. The system will display a consistent and credible “personality” at all times during a single run of the program.
11. The system will possess multiple “personalities”, with the user selecting “who” s/he wishes to talk to before beginning a conversation.
12. The system should allow creative users (e.g. specialists in the creation of learning materials or perhaps even tutors) to compose and enter new conversation scenarios. While such users will have to follow some simple ground rules to achieve this, they should not need a detailed understanding of the system’s mechanisms. Therefore, some degree of automatic coding of new utterances will be required.

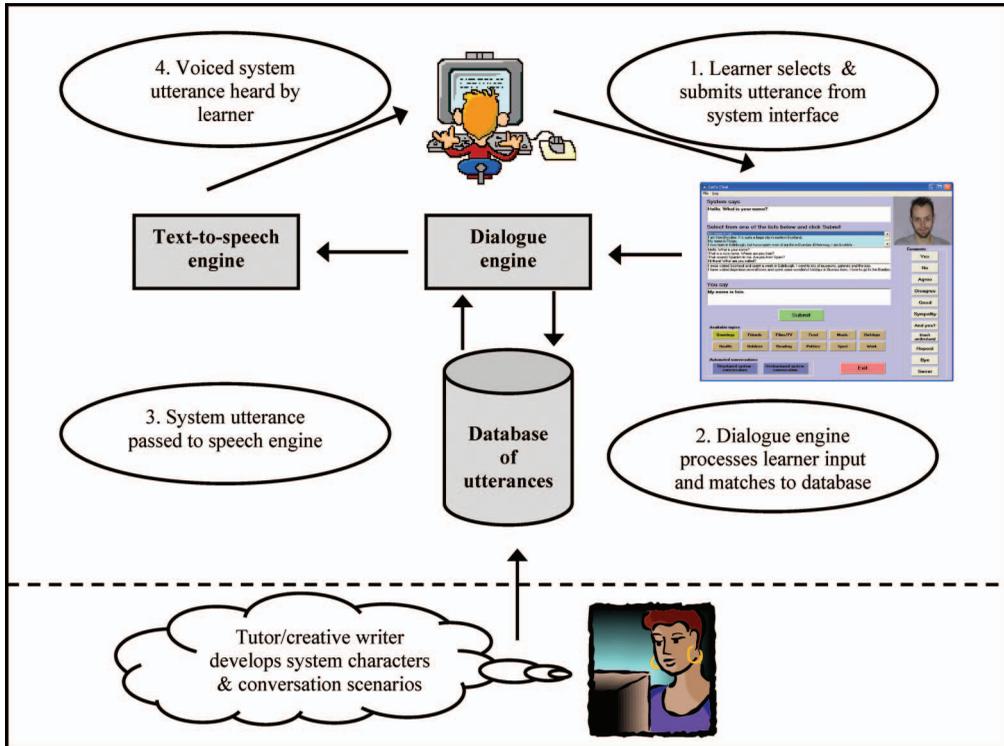


Figure 1. Illustrates the system's basic functionality

Conversation Model

Having defined the outline requirements of the proposed system, a further key decision remained to be made prior to starting the implementation process, namely the design of a conversation model. To achieve the latter, consideration was given to the characteristics of social conversation or small talk. Usually, small talk kicks off with a safe topic, such as the weather or some form of greeting exchange, and then branches either into topics relating to the shared context of the interlocutors (e.g. the party they are attending) or into a discussion of topics of personal relevance to one or both participants (Schneider, 1987). The initial moves of such communication consist typically of a question–response pair (Schneider, 1988). The way in which the conversation proceeds is open to much greater variety, but a possible progression is to a further question–response pair elaborating upon the initial exchange. In the prototype of our system, each topic-based chat concludes with a brief story. The basic conversation model, therefore, is:

Topical bid, e.g.	“Do you have any hobbies?”
Topical bid reply, e.g.	“My favourite hobby is surfing the Internet.”
Elaboration prompt, e.g.	“Why do you like the Net so much?”
Elaboration reply, e.g.	“I use chat rooms to keep in touch with my friends abroad.”
Brief story, e.g.	“I like using the Net too, but spend most of my time looking for bargains on eBay.”

After the brief story, the conversation is expected to move to a fresh topic. In this way, an interaction is built up from a series of turn-taking, topic-based sub-conversations. While the shift between topics in Let's Chat is more abrupt than would be encountered in most natural conversations, managing a gradual transition was deemed too awkward technically for the first prototype and is an issue that would be addressed in future versions of the system.

It is important to note that while the system relies on the above model when responding to an input, the user is not locked into the structure. For instance, it may on occasion be appropriate to answer a system topical bid with either the expected topical bid reply or an elaboration reply or even a brief story, and the user is free to choose any of these options.

Although we have adopted the conversation model described above for the initial prototype of Let's Chat, a worthwhile future development would be to incorporate a range of further models into the system, facilitating a greater variety of conversation structures and thus enhancing the flexibility and realism of interactions.

Comments

As indicated in system requirement 5, Let's Chat conversations can be augmented through the insertion of multipurpose comments. Available comment categories are: Yes, No, Good (e.g. “That's good”), Sympathy (e.g. “What a pity!”), Agree (e.g. “I agree”), Disagree (e.g. “I disagree”), And you? (“What about you?”), Repeat (e.g. “Could you repeat that, please?”), Don't understand (e.g. “I don't understand what you just said”), Swear (e.g. “Damn!”), and Bye (e.g. “Is that the time? I must dash”). Such comments feature prominently in natural dialogue between native speakers and can play a significant role in sustaining the conversation, so it is important to encourage their use by L2 learners. The following sub-conversation shows a variety of comments:

User topical bid: "Do you read a lot?"
 System topical bid reply: "I read a lot of magazines."
 User "good" comment: "How nice!"
 User elaboration prompt: "What magazines do you read?"
 System elaboration reply: "I like computer magazines, especially PC Plus."
 User "disagree" comment: "Can't say I agree with you."
 User brief story: "I prefer reading about really interesting things like cars and motor racing."
 System "bye" comment: "Sorry, but I have to go now. Bye."

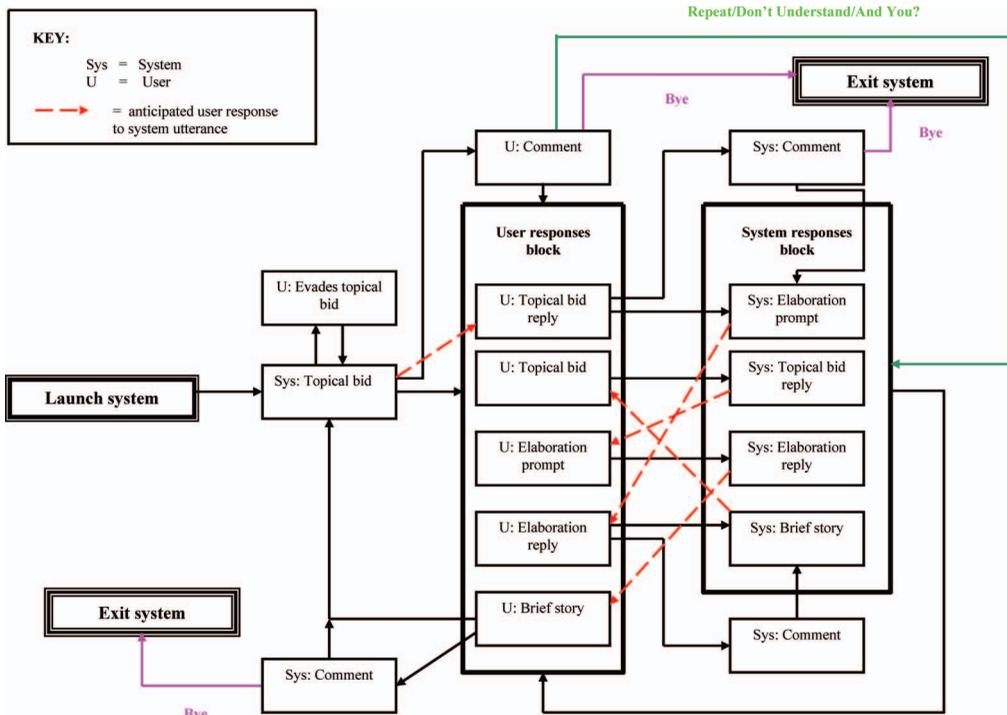


Figure 2. Illustrates all possible pathways through the conversation model

The use of comments by both system and user constitutes the only deviation from a strictly turn-taking structure, as in the current system most comment categories will be followed by another more substantive utterance from the same party.

The Let's Chat Prototype

The initial version of Let's Chat has been implemented using Microsoft Visual Basic 6, a powerful, easy-to-use programming language ideally suited to prototyping and

rapid application development. The Visual Basic dialogue engine communicates via SQL (Structured Query Language) with a Microsoft Access database containing all the system and user utterances. SQL statements are embedded into the program code as templates, the parameters of which are completed dynamically at runtime in response to the current conversation state. System utterances are voiced through a RealSpeak text-to-speech engine, originally devised by Lernout & Hauspie and now marketed by ScanSoft.

Conversation material is stored in four main tables in the database: one containing questions, the second replies, the third brief stories, and the fourth comments. All utterances within these tables can be accessed by both user and system. There is a further table holding what we term “strange reply responses”, the error alerts issued by the system when an inappropriate user input is detected. Aside from the Question and Reply tables, which are linked in a one-to-many relationship (meaning that each question has potentially more than one valid reply), little use is made at present of the relational capabilities of the Access database management system, with the other tables existing as independent entities. More intricate relational models could have been used, but would have introduced unnecessary complexity into implementation of the initial prototype. The link between questions and replies could, for instance, be modelled as a many-to-many relationship, since it is feasible that a single reply utterance could provide an adequate response to more than one question (e.g. a reply such as “I love football” could answer either “What is your favourite hobby?” or “Are you interested in sport?”).

The most significant attributes attached to utterances in the Question and Reply tables are UtteranceType (“Topical_bid” or “Elaboration”, corresponding to the phase of the conversation model in which the utterance is most likely to be used), Topic, and SubTopic. The contents of the Story table are similarly classified by Topic and SubTopic, and those of the Comment table by CommentType, indicating the category of comments to which each utterance belongs (e.g. “That’s great!” is classed as a “good” comment). Two further attributes are associated with each utterance in the Question, Reply, Story and Comment tables: Personality, indicating the system character to which that utterance is assigned, and Used, which keeps tally of the number of times the utterance has been selected as part of the system’s side of conversations.

User utterances pertaining to the active topic of conversation are arranged on the interface in three lists, each corresponding to one of the main database tables: replies, questions and brief stories. Background highlighting guides the user towards the list from which it is anticipated that the next utterance should be selected. When an utterance is selected by the user, it is displayed prominently in the “You say” text box; if the user is satisfied with the choice, s/he must then click the “Submit” button to send the utterance to the dialogue engine and initiate production of a system response.

At launch, the system takes the initiative and issues a greeting, such as “Hello. What is your name?”, which is both displayed on the screen and voiced via the speech engine. Within the Question table of the database, this utterance is defined as a

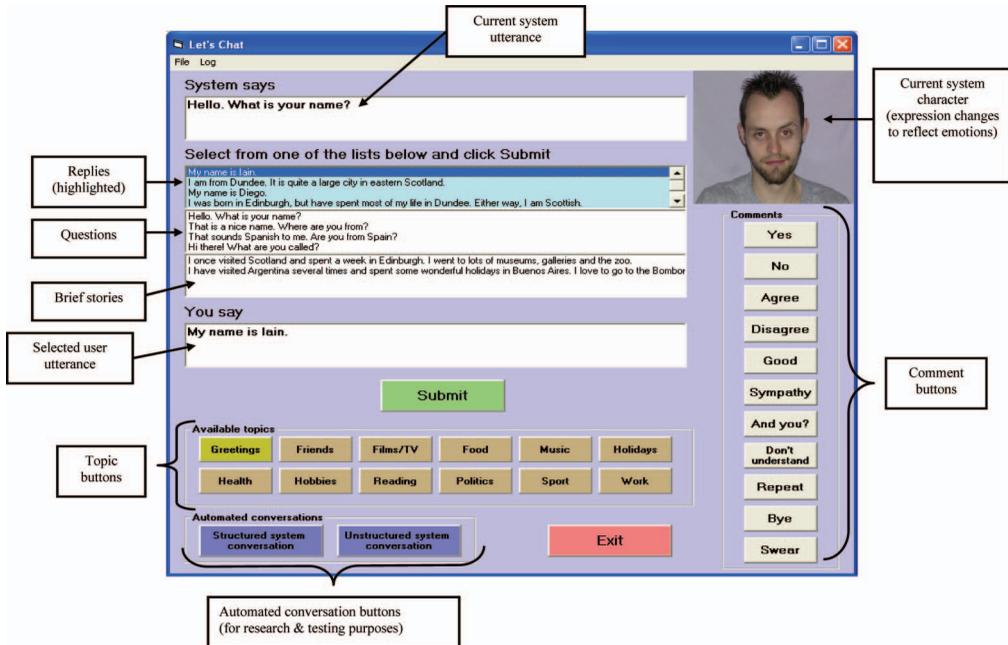


Figure 3. Shows the Let's Chat interface at launch

“Topical_bid” with the topic label “greeting”. Since topical bid questions are, by definition, the opening moves of a conversation or sub-conversation, no distinct subtopic tag can be assigned. However, to avoid the presence of nulls within the database, which can give rise to implementation problems, the topic label is duplicated in the SubTopic column. Hence, this utterance appears in the Question table as:

UtteranceText	UtteranceType	Topic	SubTopic
Hello. What is your name?	Topical_bid	greeting	greeting

While it is expected that the user will respond by selecting and submitting an appropriate reply from those displayed on the screen, s/he could choose to move the conversation directly into another area by clicking on one of the topic buttons arranged in the lower portion of the system interface and then selecting a suitable utterance (e.g. the user might choose the “Friends” button, triggering the display of utterances pertaining to that topic, and then select the question “Do you have many friends?”). Indeed, while the default setting is a system-driven conversation, the user is free to seize the initiative at all times.

To illustrate the basic mechanisms by which the system processes and responds to a user input, let us pursue the conversational scenario posited above. The user has already sidestepped the system character’s initial topical bid (“Hello. What is your name?”) and moved the conversation into the realm of friends (“Do you have many friends?”). This new topical bid is tagged in the database as:

UtteranceText	UtteranceType	Topic	SubTopic
Do you have many friends?	Topical_bid	friends	friends

↓ Dialogue engine generates SQL statement to seek least used topical bid response from Reply table pertaining to topic of friends and current system personality (in this case, we will call him Jim)

```

SELECT * FROM Reply
WHERE UtteranceType = 'Topical_bid'
AND Topic = 'friends'
AND Personality = 'Jim'
AND Used = (SELECT MIN (Used)
             FROM Reply
             WHERE UtteranceType = 'Topical_bid'
             AND Topic = 'friends'
             AND Personality = 'Jim')
    
```

↓ Query yields following system response

UtteranceText	UtteranceType	Topic	SubTopic
I have very many friends both at home and abroad.	Topical_bid	friends	friendsmany

↓ User chooses to develop current conversation topic by eliciting further information

UtteranceText	UtteranceType	Topic	SubTopic
Do you visit your foreign friends often?	Elaboration	friends	friendsmany

↓ Dialogue engine generates SQL statement to seek current system character's least used elaboration reply with same subtopic as user's question

```

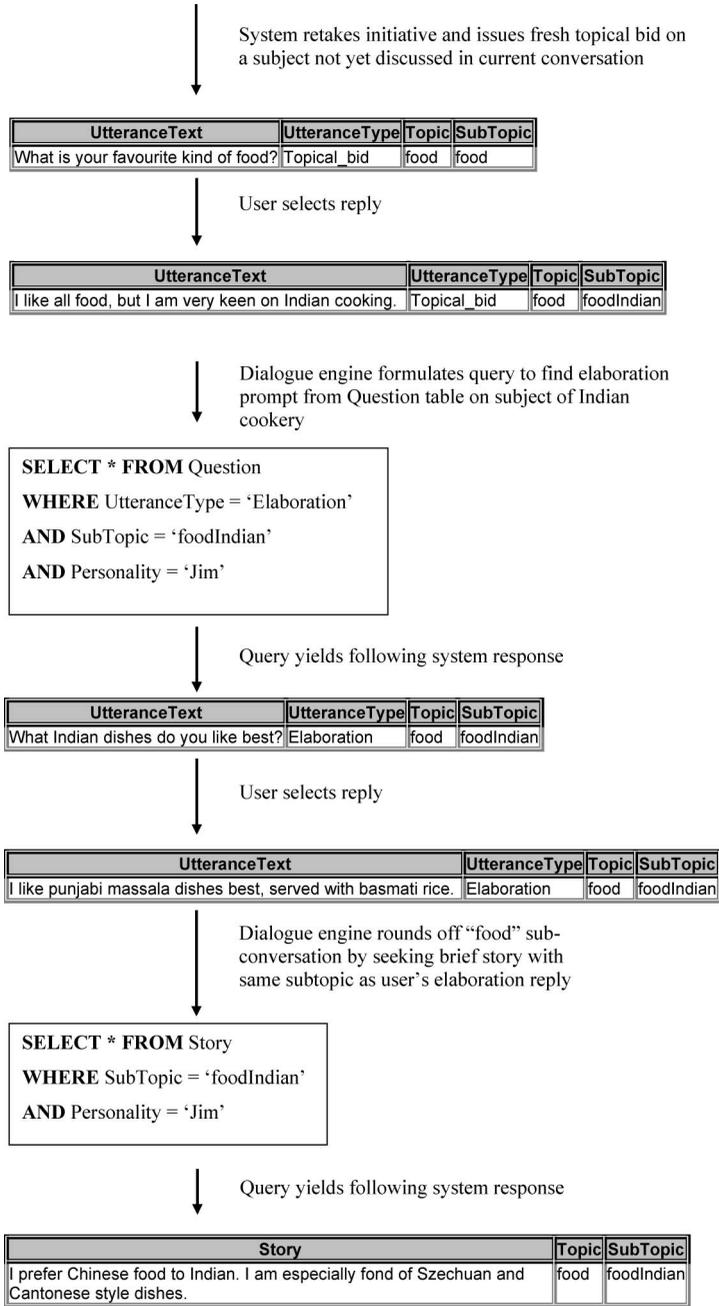
SELECT * FROM Reply
WHERE UtteranceType = 'Elaboration'
AND SubTopic = 'friendsmany'
AND Personality = 'Jim'
AND Used = (SELECT MIN (Used)
             FROM Reply
             WHERE UtteranceType = 'Elaboration'
             AND SubTopic = 'friendsmany'
             AND Personality = 'Jim')
    
```

↓ Query yields following system response

UtteranceText	UtteranceType	Topic	SubTopic
Most of my foreign friends live in South America, and I go to visit them every couple of years.	Elaboration	friends	friendsmany

↓ User rounds off "friends" sub-conversation with relevant brief story

Story	Topic	SubTopic
That must be great fun but very expensive. I once flew to Brazil and it cost around eight hundred pounds.	friends	friendsmany



The foregoing diagrammatic explanation demonstrates the basic methods by which the Let's Chat prototype sustains a conversational interaction. One significant issue remains to be mentioned, namely how the dialogue engine periodically inserts comments into the conversation. Clearly, any given user utterance can be followed

appropriately by only some of the available comment types (if, for example, the user has just stated that her pet dog had been run over last week, it would not be fitting for the system to interject “That’s good!”, but a “sympathy” comment, such as “What a shame!”, would be suitable). To enable the selection of pertinent system comments, information is stored in the database alongside each record in the Reply and Story tables about which comment categories could reasonably follow that utterance. While a “sympathy” comment would probably be best in the case of the user’s ill-fated canine, it is also conceivable that the system character could express his shock with a “swear” comment. Thus, both comment categories are marked as suitable in the database record. A similar strategy is used in the Question table to indicate if a question submitted by the user could be answered adequately by a simple yes or no.

To enhance the user experience and as an aid to comprehension of system utterances, an image representing the current Let’s Chat personality is displayed in the upper right corner of the screen at all times. A range of facial expressions reflects the character’s emotions depending on what has just been said, helping to give the system a more human dimension. A few of the available expressions for one personality are illustrated in Table 1.

In the event of the user failing to respond to a system utterance, a video clip is played, showing the system character waiting and becoming increasingly impatient, before the system eventually attempts to revive the conversation by issuing a fresh topical bid.

At present, diagnosis of user errors is enabled through a simple topic-based approach. If the user answers a question by selecting a reply or brief story belonging to a different topic, the system will come back with a “strange reply response” such as “That sounds odd. Perhaps you should choose again”. Operating on the assumption that it is more important to sustain conversational momentum, and thus hold the user’s attention, than to insist on absolute accuracy, the system will move to another question after three consecutive inadequate responses.

How best to handle user errors is an important philosophical question in a system such as Let’s Chat. In our view, a delicate balance is required, providing a degree

Table 1. Facial expressions in Let’s Chat

Happy	Bored	Shocked	“I agree”
			

of tutorial support while still allowing for a relatively free-flowing and natural interaction. Hence, we turned away from adopting an overtly tutorial manner and specifying system intervention at any user utterance that is not contextually ideal, favouring instead an approach that permits the conversation to continue except when the user's selection is fundamentally flawed. Put another way, system intervention aims to emulate the type that would be made by a sympathetic native speaker in conversation with an L2 learner. Further refinements to the error diagnosis facility may be made in the future and could include availability of different modes, thereby supporting varying degrees of tutorial intervention and feedback according to the needs of individual learners.

By default, the Let's Chat prototype generates a log of all user–system interactions in a text file, recording alongside each utterance the date and time at which it was triggered and where it fits into the conversation model. As well as proving useful for research purposes, the log could play a helpful role when learners and tutors are reviewing Let's Chat practice sessions. An edited extract from the log is included as an Appendix to this paper.

How Let's Chat would Enhance L2 Acquisition

Let's Chat could serve either as a self-access supplement to classroom activities or as a fully integrated component of an intermediate L2 curriculum. During a typical Let's Chat session, learners would participate in many short role-plays, interacting with a variety of appealing system characters and ranging over a number of the topics regularly encountered in native social conversation. By rehearsing dialogues with the computer, it is anticipated that learners would begin to assimilate the pre-stored utterances of which the interactions are composed, acquire insights into the skill of conversation management in the target language and gradually build confidence in their communicative capabilities, all of which could be carried forward into subsequent natural language interactions.

For a system such as Let's Chat to facilitate the achievement of its learning outcomes, it is, of course, crucially important to capture and hold learners' attention. To this end, we have focused not only on producing a system that could enable good quality interactions between learner and computer, but also on rendering the overall user experience as entertaining as possible.

The nature of the conversation material entered into the Let's Chat database contributes significantly to the system's appeal. While the prototype's database has been populated by the main developer of the system, our aim is to work jointly with creative writers and specialists in the development of language learning materials to prepare future versions, thereby producing content that is both entertaining and specifically appropriate for use in an L2 environment. In conjunction with the visual stimulus provided by the lively facial expressions of the selected personality, good conversation material should make practice with Let's Chat an enjoyable experience conducive to successful L2 acquisition.

Future Developments

In the course of developing the Let's Chat prototype, a number of areas for future work have been identified, three of which have been mentioned already (greater flexibility of conversation models, enhanced response to user errors and creation of appropriate and entertaining content). Others of significance include:

1. permitting users to introduce limited modifications to pre-stored utterances to reflect their own communicative requirements;
2. development of a structured interface to ease input of new conversation material by creative users or tutors (see system requirement 12); and
3. use of real speech recordings for playback of system utterances.

Although a pre-stored utterance dialogue system functions on the basis that all possible user inputs are known, it should be possible to permit limited modification of the available phrases and still maintain a coherent interaction with sensible system responses. One way to do this is to introduce "slots" into appropriate pre-stored utterances, which are filled by selecting from a dropdown list. For instance, the user could complete the stem phrase "My favourite hobby is . . ." by choosing the option most relevant to his/her own preferences from a catalogue of possible hobbies. Managing the system response under such circumstances requires the database either to contain a specific follow-up utterance for every user option or to include a number of more generic responses appropriate for the majority of cases. For example, if the user completes the "My favourite hobby is . . ." stem phrase with the "playing golf" slot, the system could respond in a tightly-linked fashion by enquiring "What is your handicap?" or take the generic route and ask simply "Why do you like it?". Probably the most practical approach is a blend of the two strategies: storing precise follow-ups where they appear necessary and adopting the broad-brush option when it is feasible to do so.

Development of a structured interface for the input of new conversation material is an important step towards making Let's Chat an "open" tutoring tool that can be tailored to the requirements of different groups of learners. Since our aim is to allow creative users to generate fresh conversation scenarios without understanding the techniques employed to classify utterances within the system database, an automatic coding facility would be necessary to ensure that new phrases are correctly stored. One way to achieve this is to provide a form-filling type of interface based on the current system personality and active conversation model. Taking the prototype's conversation model as an example, at least one topical bid, topical bid reply, elaboration prompt, elaboration reply and brief story would have to be entered into corresponding text boxes on the interface for each new conversation scenario. When satisfied with the material and after suggesting an appropriate topic (and, in some cases, subtopic label) for the conversation, the creative user would simply click a "Submit" button to generate the relevant "metadata", or coding, for the utterances and consign them to the system database. While this is quite technically demanding, it is achievable.

Using real speech recordings for voicing system utterances in place of a text-to-speech engine offers a number of potential benefits. First, it is recognized that even the best speech synthesizers currently available are unable to cope at all times with the manifold complexities of pronunciation and prosody (see, for example, McTear, 2002, p. 127). Inevitably, the occasional utterance is intoned in a way that sounds curious to a native speaker. Although such deficiencies are not usually severe enough to impede understanding, they are clearly undesirable within a system designed to promote good L2 acquisition. It would be a relatively simple, if somewhat laborious, task to pre-record all utterances stored in the system database and hence circumvent most problems of this nature. It should be noted, however, that even recording does not always guarantee perfect prosody, as concatenation of utterances can lead to complications. Within the context of the Let's Chat prototype, it is possible to foresee problems when certain system comments are followed immediately by another utterance.

A further potential benefit of using recorded speech is that system characters could each possess their own accent, thereby exposing learners to one element of the variety of target language they are likely to encounter in everyday life. The difficulty of coming to terms with variations of accent is, of course, frequently reported by L2 learners when travelling to a country in which the target language is spoken, so any way of preparing would surely be advantageous. Indeed, the prospect of utilizing Let's Chat to prepare for the diversity of regional accents was warmly welcomed by members of one potential user group (Chinese-speaking students living in Scotland) during interviews with the developer.

There is, however, one disadvantage of using recorded speech in the Let's Chat system, namely that it would complicate the process of generating new conversation material for creative users and tutors by adding the burden of recording every new phrase in an appropriate voice for the system personality. That said, we believe that the dividends of using real speech recordings are sufficiently high to warrant inclusion of the facility to do so. Perhaps, therefore, the answer lies in equipping the system with two modes, one in which all utterances are voiced via recorded speech, the other using only text-to-speech.

Conclusions

Building upon theoretical work in the field of applied linguistics that suggests the importance of formulaic language in SLA, this paper has proposed that a human-computer dialogue system based on communication by pre-stored utterances could offer a fertile environment for the acquisition and rehearsal of L2 social conversation skills. A description of Let's Chat, a prototype under development at the University of Abertay Dundee, is provided. It is argued that the quality of interactions mediated by this system is such that learners could reasonably expect to enhance their grasp of idiomatic, native-like modes of expression by "conversing" with it, thereby achieving higher levels of confidence and fluency in subsequent natural language interactions with human partners. Future work will focus primarily on rendering Let's Chat interactions more flexible and true to life.

Acknowledgements

We wish to thank the Carnegie Trust for the Universities of Scotland Larger Grants Scheme for its generous support of the project “A Flexible Dialogue System for Second Language Learning”, and our co-participants from the University of Abertay Dundee and other Scottish higher education institutions: Alison Cawsey (Heriot-Watt University), Carmen García del Río (University of St Andrews), Petra Leimich (Abertay), Robin Lickley (Queen Margaret University College), Astrid Schepman (Abertay). We are also grateful to Matt Rowley for his contributions to the visual dimension of Let's Chat and to Margaret Ross and Jenny Reid for their helpful comments on earlier drafts of this paper.

References

- Charniak, E. (1995). Natural language learning. *ACM Computing Surveys*, 27(3), 317–319.
- Feigenbaum, E. A. (2003). Some challenges and grand challenges for computational intelligence. *Journal of the ACM*, 50(1), 32–40.
- Fillmore, C. J. (1976). The need for a frame semantics in linguistics. In *Statistical methods in linguistics*. Stockholm: Skriptor.
- Lewis Johnson, W., Marsella, S., Mote, N., Viljálmsón, H., Narayanan, S., & Choi, S. (2004). Tactical language training system: Supporting the rapid acquisition of foreign language and cultural skills. In *Proceedings of InSTIL/ICALL2004—NLP and Speech Technologies in Advanced Language Learning Systems*, Venice, 17–19 June.
- McTear, M. F. (2002). Spoken dialogue technology: Enabling the conversational user interface. *ACM Computing Surveys*, 34(1), 90–169.
- Schneider, K. P. (1987). Topic selection in phatic communication. *Multilingua*, 6(3), 247–256.
- Schneider, K. P. (1988). *Small talk: Analysing phatic discourse*. Marburg: Hitzeroth.
- Stewart, I. A. D. (2004). Adapting the Contact System for Spanish Language Learning. Unpublished MSc dissertation, University of Abertay Dundee.
- Todman, J. B., & Dugard, P. (2001). *Single-case and small-n experimental designs*. Mahwah, NJ & London: Lawrence Erlbaum.
- Vihman, M. M. (1982). Formulas in first and second language acquisition. In L. K. Obler & L. Menn (Eds.), *Exceptional language and linguistics* (pp. 261–284). New York: Academic Press.
- Vlugter, P., Knott, A., & Weatherall, V. (2004). A human–machine dialogue system for CALL. In *Proceedings of InSTIL/ICALL2004—NLP and Speech Technologies in Advanced Language Learning Systems*, Venice, 17–19 June.
- Weinert, R. (1995). The role of formulaic language in second language acquisition. *Applied Linguistics*, 16, 180–205.
- Wilks, Y., & Catizone, R. (2001). Human–computer conversation. In A. Kent (Ed.), *Encyclopedia of library and information science*. Vol. 69. New York: Dekker.
- Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, 18, 47–67.
- Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching*, 32(1), 213–231.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principles and practice. *Applied Linguistics*, 21(4), 463–489.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Appendix: Transcript of a Let's Chat conversation

Sys: Initial topical bid Hello. What is your name?
 User: Topical bid reply My name is Diego.
 Sys: Elaboration prompt That sounds Spanish to me. Are you from Spain?
 User: Elaboration reply I come from Buenos Aires, the capital of Argentina.
 Sys: "Good" comment How nice!
 Sys: Brief story I have visited Argentina several times and spent some wonderful holidays in Buenos Aires. I love to go to the Bombonera and watch a Boca – River game.

 User: "Agree" comment I couldn't agree more.
 Sys: New topical bid Where do you go on holiday?
 User: Topical bid reply I like to go to the mountains for my holidays.
 Sys: Elaboration prompt Why do you go to the mountains?
 User: Elaboration reply I enjoy climbing.
 Sys: "Agree" comment I share your view.
 Sys: Brief story I am very keen on climbing. I have visited the Alps several times and once reached the summit of the Matterhorn.

 User: "Swear" comment Bloody hell!
 Sys: Comment response There's no need to use bad language!
 Sys: New topical bid I enjoy watching films. Do you?
 User: Topical bid reply I love films, especially Westerns.
 Sys: Elaboration prompt What is your favourite film?
 User: "And you?" comment What about you?
 Sys: Elaboration reply My favourite film is Fort Apache.
 User: Makes topical bid What sort of music do you listen to?
 Sys: Topical bid reply I listen to lots of music, but mainly pop.
 User: Elaboration reply I like rap and hip-hop. I am a big fan of Eminem.
 Sys: Brief story I like old-fashioned rock better than rap. Give me Bruce Springsteen any day. Born in the USA is one of my favourite albums.

 Sys: New topical bid Do you read a lot?
 User: "Repeat" comment Would you mind repeating that, please?
 Sys: Repeat topical bid Do you read a lot?
 User: "Yes" comment Yes.
 Sys: Elaboration prompt Who is your favourite author?
 User: Elaboration reply John Grisham is a terrific writer. I could read his books all day.

 Sys: "Good" comment That's nice.
 Sys: Brief story I like Grisham too, but my favourite crime writer is Jonathan Kellerman.

 User: "Bye" comment Is that the time? I've got to go now.
 Sys: Comment response OK, see you later!