*Natural Language Engineering* **1** (1): 1 – 12. Printed in the United Kingdom 


# *User-Guided System Development in Interactive Spoken Language Education*

E. ATWELL, University of Leeds, Yorkshire, UK
*P. BALDO, DIDA*EL S.r.l., Milan, Italy*
*R. BISIANI, Universita' di Milano Bicocca, Italy*
*P. BONAVENTURA, Universitaet Hamburg, Germany*
*D. HERRON, Universitaet Hamburg, Germany*
*P. HOWARTH, University of Leeds, Yorkshire, UK*
*W. MENZEL, Universitaet Hamburg, Germany*
*R. MORTON, Entropic Cambridge Research Laboratory Ltd, Cambridge, UK*
*C. SOUTER, University of Leeds, Yorkshire, UK*
*H. WICK, Ernst Klett Verlag, Stuttgart, Germany*

(*Received 22 Nov 1999; revised 29 February 2000*)

**Abstract**

This paper is a Case Study of user involvement in the requirements specification for project ISLE: Interactive Spoken Language Education. Developers of Spoken Language Dialogue Systems should involve users from the outset, particularly if the aim is to develop novel solutions for a generic target application area or market. As well as target end-users, SLDS developers should identify and consult "meta-level" domain experts with expertise in human-to-human dialogue in the target domain. In our case, English language teachers and publishers provided generic knowledge of learners' dialogue preferences; other applications have analogous domain language experts. These domain language experts can help to pin down a domain-specific sublanguage which fits the constraints of current speech recognition technology: linguistically-naive end-users may expect unconstrained conversational English, but in practice dialogue interactions have to be constrained in vocabulary and syntax. User consultation also highlighted a need to consider how to integrate speech input and output with other modes of interaction and processing; in our case the input speech signal is processed by speech recogniser, stress and mispronunciation detectors, and output responses are text and graphics as well as speech. This suggests a need to revisit the definition of "dialogue": other SLDS developers should also consider the merits of multimodality as an adjunct to pure spoken language dialogue, particularly given that current systems are not capable of accurately handling unconstrained English.

## 1 Introduction

This paper presents, as a case study, some lessons we have learnt on project ISLE: Interactive Spoken Language Education. [1] The design and development of our Spo-

[1] This research was supported by the European Commission's research and development programme (DG XIII), under its Framework 4 Language Engineering project ISLE LE4-8353, see *http://nats-www.informatik.uni-hamburg.de/isle/*

ken Language Dialogue System was guided from the outset by consideration of user requirements; for us, user requirements were at least as important as technical constraints in SLDS engineering. One reason for this was that, although our project workplan included technical workpackages and goals, the target dialogue types and tasks did not seem as clearly defined as for some other SLDS applications. The strategic goal of ISLE was to exploit available speech recognition technology to improve the performance of traditional language learning software. The project was targeted at adult German and Italian learners of English. The project developed a number of specialised speech recognition components, outlined below. A complementary research strand was a detailed survey and analysis of prospective user requirements.

## 2 Guided by end-users and meta-level domain experts

We sought expert advice and opinions from a range of prospective end-users (learners of English as a second language), as well as professionals and practitioners in English language teaching (ELT teachers and researchers), and industry experts in the ELT market (publishers of ELT resources, textbooks and multimedia). The ISLE project partners included representative users: English language learners at all 6 sites [2]; Leeds University is a centre of excellence in English language teaching and research; Klett is a major German publisher of ELT resources and textbooks; and Dida*el is a major Italian publisher of multimedia educational systems.

At the outset of the Project, we sought guidance on what users need in a computer system for pronunciation tutoring in English as a foreign language. The conclusions drawn from this analysis of user requirements led us to our solution: specifications of the ISLE system structure, exercise types, ISLE calls available to the high-level interface, and user-centred Corpus-based verification and validation.

We believe that an important principle in Spoken Language Dialogue Systems (SLDS) development is to involve users at every stage of development and evaluation. Other researchers in Human Computer Interaction have advocated this *e.g.* (Dix *et al.* 1998); we believe it is particularly important when aiming to promote takeup of ground-breaking technology, such as SLDS. Otherwise, there is a danger that research and development will focus on theory rather than prospective market demands; and without user consultation, potential users may remain unaware of the benefits offered by SLDS:

"In general, UK industry seemed to consider that too much effort had been spent in the past on esoteric academic problems ... Another problem was the lack of awareness of speech and language technology amongst potential users. It was also felt that there was a lack of useful, analysable feedback." (EPSRC 1998)

"On the whole, developers have not asked users what they want from speech and language technology. Research has been driven by the interests of researchers rather than

[2] ISLE project consortium: Dida*el S.r.l.(Milan, Italy), Entropic Cambridge Research Laboratory Ltd. (Cambridge, UK), Ernst Klett Verlag (Stuttgart, Germany), University of Hamburg (Germany), University of Leeds (UK), University of Milan Bicocca (Italy).

potential customers, and has tended to focus on tackling theoretically interesting problems rather than examining usefulness. In the past, for example, much computational linguistics research was led by linguists whose primary aim was to implement linguistic theories" (Atwell 1999)

There are two distinct classes of user who should be consulted:

- Prospective end-users, who will be speaking to and listening to the eventual SLDS (if they are persuaded to use it); they can advise on individual requirements and preferences;
- Meta-level experts: practitioners and professionals whose service the SLDS will be replacing or augmenting, who can contribute a more generic insight into end-user practices and preferences; and Industry experts from the companies which aim to integrate the SLDS into their current systems and sell to the end-users, who can provide an insight into end-users as customers.

It is not enough to involve end-users in evaluation trials of the prototype SLDS application. At this stage, it is too late for users to point out functionality missing from the design, let alone question whether Spoken Language Dialogues are the best and/or only solution to the given task; and it may be hard to convince the eventual market that user involvement is more than an afterthought. One approach to greater involvement of prospective users is to set up an expert advisory panel representing users and meta-level experts, and consult this at various stages throughout the project; for example, such panels have advised the British National Corpus (BNC, (Leech 1992)) and Multilevel Annotation Tools Engineering (MATE, (Dybkjaer *et al.* 2000)) projects.

We advocate going one step further, that user groups should be represented by full partners in SLDS development projects. We believe that other developers of Spoken Language Dialogue Systems should involve both end-users and meta-level experts from the outset, particularly if the aim is to develop novel solutions for a generic target application area or market. It might appear that not many target application areas for dialogue systems have comparable generic meta-level experts: as English language learning is the subject of significant academic theoretical study, there are many more academic experts to consult on the language of ELT dialogues than, say, the language of Telesales, market research interviews, or Air Traffic Control dialogues. However, expertise on dialogue structures is not confined to academic linguists. For example:

- The use of pre-planned, structured interview dialogues within focus groups of potential customers is a standard technique in advertising and market research; this yields qualitative results which contrast with and complement quantitative research results such as those gleaned from multiple-choice questionnaires. Market research industry experts have significant knowledge of the organisation and structure of qualitative market research dialogues, and can provide meta-level insight to guide projects in this domain, *e.g.* see (Wilson and Rayson 1993).
- Pilots and Air Traffic Controllers can advise SLDS projects in this domain,

*e.g.* (Churcher *et al.* 1996), (Churcher *et al.* 1997), as both prospective end-users and as meta-level experts, as they have specialised training and experience of the controlled language of ATC dialogues, as prescribed by aviation authorities (CAA 1992).

- Telesales providers have considerable knowledge of dialogue types and structures used in their operations, which can inform SLDS design; for example, telesales operators for Virgin Rail and GE Capital are trained to follow generic dialogue scripts according to customer/enquiry type (Meekings 1999), (Brett 1999).

## 3 Different types of "dialogue"

SLDS designers, end-users and meta-level experts may have different perspectives on what counts as "dialogue", and what kinds of interaction to expect from a SLDS. The Collins English Dictionary (Collins 1994) definition is:

*dialogue: 1. conversation between two or more people. 2. an exchange of opinions on a particular subject; discussion. 3. the lines spoken by characters in drama or fiction ...*

CED goes on to give 5 further (more specialised) senses. The significant point for SLSD design is that linguistically-naive end-users may expect unconstrained conversational English. In practice, SLDS dialogue interactions have to be tightly constrained in vocabulary and grammar, to enable current speech recognisers to perform with acceptable accuracy levels. Fortunately, domain language experts can help to pin down a domain-specific sublanguage which fits the constraints of current speech recognition technology. For the three example applications cited above (market research interviews, Air Traffic Control, and Telesales), domain language experts can provide insight to guide the design of a sublanguage the speech recogniser can cope with. In our case, English language learners, the end-users, tended to expect general conversation practice, with minimal limitations on vocabulary and syntax (particularly challenging for the speech recogniser given that it would have to cope with learners' errors). However, the domain experts, English language teachers and publishers, were able to suggest a range of interactions or exercise types which strictly constrained the user's input while still involving the user in a learning dialogue.

User consultation highlighted a need to consider how to integrate speech input and output with other modes of interaction and processing; in our case the input speech signal is processed by speech recogniser, stress and mispronunciation detectors in parallel, and output responses involve text and graphics as well as speech. This suggests a need to revisit the definition of "dialogue", to include other modes of interaction between human and computer. Other SLDS developers are also considering the merits of multimodality as an adjunct to pure spoken language dialogue, particularly given that current systems are not capable of accurately handling unconstrained spoken English; for example, see (Wyard and Churcher 2000).

## 4 Analysis of ISLE user requirements

The major issues anticipated from the investigation of user needs were:

- what analysis is expected or required of the users' spoken input, in addition to speech recognition
- the speech, text and visual data to be presented to users as the system's contribution to the dialogue
- the activities that users will engage in, including other forms of interaction as well as spoken dialogues

The major sources of meta-level expert advice were:

- consultation of English language teaching professionals (teachers, researchers, publishers), by informal discusions and questionnaires distributed via a range of channels:
  - — ISLE project participants
  - — Staff of the Language Centre, University of Leeds
  - — BALEAP members (British Association of Lecturers in English for Academic Purposes)
  - — EUROCALL subscribers (the European Association for Computer Assisted Language Learning) e-list
  - — TESLCALL subscribers (Teachers of English as a Second Language CALL) e-list
  - — ICAME (International Computer Archive of Modern and medieval English) annual conference participants
  - — NLP-CALL (Natural Language Processing for Computer Assisted Language Learning) workshop participants
  - — IATEFL (International Association of Teachers of English as a Foreign Language) Pronunciation special interest group
  - — CAPITAL (Computer Assisted Pronunciation Investigation Teaching and Learning) special interest group
  - — CALICO (Computer Assisted Language Instruction Consortium) e-list
- Evaluation by language learners and professionals of existing software
- A survey of the literature on pronunciation teaching

We also surveyed end-user preferences, via a questionnaire administered to adult learners of English. Details of the advice gathering exercise are in (Atwell *et al.* 1999). This survey of linguists' requirements guided the subsequent development of our ISLE system; as well as its general relevance to SLDS user preferences, it should also be of specific relevance to other researchers integrating speech recognition into language-learning systems, for example (Dalby *et al.* 1998), (Eskenazi and Hansma 1998), (Witt and Young 1997). The main analysis findings are summarised below:

- The units of speech presented and practised in the dialogue between system and learner-user should include:
  - — single syllables for ear training, perhaps reinforced with video production of problem phonemes, again possibly reinforced visually;

- — single words for phonemes and word stress
- — phrases for stress patterns and weak forms
- — sentences for contrastive intonation
- — utterances presented with wave forms, orthographic text aligned with the graphics and with some kind of indicator moving along the wave form as the sound is heard (e.g. by the colour changing with the speech)
- — texts or scripts to be read aloud
- — conversations with turn-taking

- In all these activities, the facility should be provided for users to listen again to their own speech for comparison with the target.
- One lesson from the evaluation of Auralang is the importance of giving the user the option of active control over the recording phase (i.e. by having a start and stop button). This would clearly have implications for analysis and feedback. Control could also be provided in terms of choosing between accents to listen to, having a choice between scored and unscored practice, size of pronunciation unit to practice, etc.
- There are two views on providing as much variety as technically possible. On the one hand, this may be wasted effort if learners don't make full use of the features offered and they might be confused by a complex structure. On the other hand, variety would allow for the diversity of users' preferred learning styles and may encourage repeated use of the program. The ISLE project decided that maximum variety is desirable, with the obvious caveat of technical feasibility and cost.
- Dialogue exercises could encourage learners to activate their speech, and build confidence; a low "acceptability" setting for the recogniser would allow unstressful production with positive feedback in the form of encouraging conversational responses.
- Of the categories of system output or responses (text, sound and graphics) there is a clear preference for sound (listen and repeat etc), some interest in textual output/prompts and reservations about the value of wave forms in graphical output. Some respondents suggested that learners would have difficulty interpreting such visual information, though others thought that animated representations of articulation could be useful. The commonest additional suggestion was for learners to participate in unconstrained conversation. This was not included in the list of possible preferences to choose from, as we had already ruled this as beyond the scope of ISLE on grounds of technical feasibility; however, there is clear user demand, which could perhaps be addressed in future research.
- The survey showed a clear desire for IPA (or some version of it) to be available as an option in feedback, with some training or guide to its use as a reference tool.
- The "benchmark" representative existing system we tried out, Auralang, offered limited pronunciation guidance. Users were given a raw score (0-7) and shown a target waveform to compare against their own speech waveform; but users are given no further guidance as to how they should modify specific

aspects of articulation to produce a better match to the model. Our survey indicated that effective feedback involves a range of alternative guidance, targeted at specific linguistic errors.

- Users should be able to hear their own speech in comparison with the target and to be given guidance in analysing the difference. This may be done by means of wave forms, in which case they must be accurately aligned. The system should direct users to the significant differences and as far as possible inform them of the type of mismatch, especially the length and quality of individual phonemes.
- A context-sensitive help system could provide additional detailed guidance, perhaps with examples of correct articulation, stress pattern etc. The users' attempts at imitating the target sounds should be stored during a session and evaluated so that they can be reviewed at ant time to compare levels of success.
- A clear wish for some sort of scoring was reported, and several teachers commented on its motivating effect for learners. Some respondents had reservations about the problem of "correctness" in pronunciation scoring, and said that an indication of "comprehensibility" would be preferred. As much additional information and guidance as possible should be available, partly to allow the learner's route through the material to be individualised and varied. This would help to make repeated use of the program more stimulating.
- Teachers expressed a marginal preference for instructions and feedback in English over L1 (the student's first language) or both languages. Learners showed no interest in the use of the mother tongue on screen. Comments suggest that a choice might be offered and that the more advanced learner would choose English.
- The speech data presented should partly be selected to give German and Italian learners practice in their specific problems.
- Pronunciation training depends to a large extent on accurate, guided listening. It would be beneficial, therefore, if material were included for the introduction and training in key problem areas: the articulation of individual sounds, word stress, the contrast between strong and weak syllables, weak forms, phrase and sentence stress, and intonation.
- Activities should relate to the specific pronunciation problems which learners and teachers consider to be most important. Features related to stress received great emphasis: word stress, weak forms and intonation. Non-native teachers, and learners, place greater weight on problems related to individual sounds than do native-speaker teachers. Prominent among the further comments was the view that pronunciation of individual sounds should be presented contrastively with L1. For German learners of English the major problems are intonation, followed by individual sounds and then stress-related features. For Italians, individual sounds are the biggest problem followed by word stress.
- The user questionnaire and the literature survey both produced a wide diversity of strongly held views on the English accent learners should hear, ranging from 'anything but RP' to 'RP is a model appropriate for international in-

telligibility' to 'doesn't matter' (RP = Received Pronunciation, traditionally taken as a standard British English accent). There was an overall preference (even if sometimes reluctant) for something like standard British English as the model for users to aim at, based on educated varieties of southern British English, and possibly also northern and Scottish English. Speech presented to learners should ideally be varied by sex, age, region, with as much choice as possible offered to the user, and information about the speakers.

## 5 Meeting the requirements: outline of ISLE technical specification

The above analysis of requirements was central to developing specifications of ISLE system architecture and components, exercise types, feedback, and corpus-based verification and validation.

### *5.1 System architecture and components*

The speech processing research efforts of the ISLE project have concentrated on development in four main areas (see (Herron *et al.* 1999)): (1) an HMM-based recognition engine that is fast, reliable, and robust even when presented with non-native speech; (2) a localisation mechanism that reliably pinpoints deviations in the student's pronunciation from the native model; (3) a diagnosis mechanism that explains pronunciation errors as the product of phonemic or orthographic errors by the student, allowing the system to focus on particular errors and to provide useful feedback; and (4) a lexical- stress detector to locate misplaced stress, which is clearly a problem for many learners of English. Exercises and feedback are presented within a high- level multimedia authoring system. Models adapted to non-native speech are used during the recognition stage in order to increase accuracy. When localising and diagnosing errors, however, non- adapted models are used, to maximise the probability of detecting errors. Additionally, although any phone-level error (substitution, deletion, or insertion) can potentially be detected, the system concentrates on errors that are expected based on the student's first language or that stem from common orthographically-induced mispronunciations; this further increases accuracy and lowers the probability of false alarms. Integration of these components creates a natural learning environment in which the student is never responsible for self-diagnosis. Besides providing the student with immediate feedback, long-term performance data (at the exercise, phrase, word, and phone levels) is collected to allow the student's performance to be tracked across time.

A prototype system for individualised training has been developed, available in two versions: English pronunciation for German learners and English pronunciation for Italian learners. In the prototype, the student can interact with the system by reading a given text, by matching items, through multiple- choice exercises, and through exercises on the pronunciation of minimal pairs (e.g. dip and deep). The system evaluates the input and gives feedback about the location and type of mispronunciation, and about how it could be improved. The feedback is presented in several different forms: scoring, playback and textual transcription of an utterance, presentation of model pronunciations, highlighting of mispronunciations in the transcription, giving a textual description of the error.

The ISLE diagnostic components are capable of determining when, where, and which pronunciation mistakes have been made. ISLE also introduces a feedback module that effectively informs the end-users of what their pronunciation mistakes are and how they can compensate for or correct them. Developers can integrate the full capabilities of the ISLE system directly into their programs using authoring systems such as Asymetrix's ToolBook and Macromedia's Authorware or development environments like Microsoft's Visual Basic and Visual C++ because the system is based on the Component Object Model (COM) architecture under Win32. Underlying the prototype end-user system is a software toolkit with components for spoken language interaction in controlled exercise environments, speech quality assessment, and feedback presentation; these tools can in principle be straightforwardly interfaced to other authoring systems, via the OCX-based mid-level interface, which handles all calls to the recognition and diagnostic engines.

### *5.2 Exercise Types*

Our choice of exercise types is a compromise between user preferences uncovered in our analysis of user requirements, and what is technically feasible in the prototype system as specified above. The prototype system accepts spoken input in a controlled environment with the following exercise types:

- Reading a given text (presented as part of a simulated conversation)
- Matching items (oral combination of suitable items from several lists, such as choosing the subject of a sentence, the verb, and an object to form "she drinks coffee" or "he plays violin")
- Multiple choice exercises (oral selection form a list of different items)
- Pronunciation of minimal pairs: dip vs. deep
- Answering simple questions (e.g., "what is the girl drinking" with a picture showing that "she is drinking soda")
- Producing simple scene descriptions (e.g., "there is a house with a swimming pool")

We have estimated the active target vocabulary to between 100 and 200 words and a perplexity of 5 to 10. However the final decision will be taken after further evaluation of system perfomance in field trials.

The system evaluates the quality of the speech input and derives feedback information with respect to

- Position of a mispronunciation
- Kind of mispronunciation (phone quality, word stress patterns)
- Possibilities and directions for improvement

There are two versions of the prototype system:

1. English pronunciation for Italian learners
2. English pronunciation for German learners

### 5.3 Feedback

Effective feedback involves, among other things:

- Users can hear their own speech in comparison with the target and are given guidance in analysing the difference. The system directs users to the significant differences and as far as possible informs them of the type of mismatch, in terms of the length and quality of individual phonemes.
- A context-sensitive help system provides additional detailed guidance, with examples of correct articulation, stress pattern etc. The ISLE demonstrator uses contextual menus: the string uttered by the student is shown on screen, with mispronounced words or areas clearly marked. A contextual menu is attached to each word, allowing the user to choose from a variety of kinds of feedback.

For phone-type errors, the student can:

- Hear the word (or a region containing the word) as they said it
- Compare it to the correct pronunciation
- Receive feedback telling them what they did wrong
- See tips on how to improve their production of that phone
- Jump to exercises specific to that phone

For stress errors, the students can:

- Hear the word as they said it
- Compare it to the correct pronunciation
- Receive feedback telling them what they did wrong

The user's attempts at imitating the target sounds are stored during a session and evaluated so that they can be reviewed at any time to compare levels of success.

### 5.4 Corpus-based Verification and Validation

The requirements section suggested large-scale off-line verification, using a test corpus collected for the ISLE project; and limited on-line verification or user trials.

A medium-sized corpus of learners' spoken English has been collected, orthographically and phonetically transcribed, and annotated for phone- and stress-errors, (Souter *et al.* 1999), (Menzel *et al.* 2000); this is being used for performance evaluation. The corpus comprises speech data from 23 Italian and 23 German intermediate and advanced learners of English, balanced roughly on sex, age, and proficiency. Approximately 250 utterances were recorded by each speaker; speakers required between 20 minutes and one hour to record their contributions. The non-native speech corpus is used to optimize the recognition and adaptation parameters for non-native speech and low-perplexity recognition tasks, and to evaluate the diagnosis of mispronunciations expected from intermediate learners of English. The corpus therefore contains a representative sample of the target non-native accents and exercise types to be found in the final ISLE system.

The on-line testing was designed to be stand-alone (not tied to courseware) and completed at University of Leeds (with English language teachers and researchers) and Klett and University of Milan (with English language students). Ideally we would have liked to have pretest-posttest differences to show, but due to time and finance constraints this was not possible.

EAGLES and other Language Engineering projects have advocated user-centred evaluation (EAGLES 1995), which involves attempting to objectively assess how well the final system matches the user requirements specification. In addition to the above on-line testing, we have revisited our requirements analysis in a user-centred evaluation exercise, to check that the requirements listed in section 4 have been met.

## 6 Conclusions

Project ISLE aimed to enhance use of speech recognition technology in future Computer- Assisted Language Learning (CALL) products for adult learners of English. This paper describes ISLE user requirements and proposed solutions. The requirements of prospective users of ISLE were analysed by consulting users, that is, English language learners, teachers, and current market providers; and through a survey of the market and competitor products. This requirements analysis provided ISLE partners with guidance on what users need in a computer system for pronunciation tutoring in EFL. The main tool in this exercise has been a variety of data-gathering exercises (questionnaire surveys, software evaluation and literature survey), with interpretation and analysis. The main conclusions relate to the model of spoken English the system should use, the kinds of learning activity it should provide and the type of feedback it should aim to give. We were further guided by an analysis of the market for educational software in general and ISLE-like systems in particular, including a survey of competitors.

The solutions proposed by the ISLE project are outlined above. The ISLE system differs from previous systems because it (a) takes advantage of state-of-the-art recognition technologies that were obviously unavailable previously and (b) adds diagnosis and feedback at a much more specific level. The resulting system will be of value to ELT, and will have more credibility in the target market because ELT professionals were actively involved in design.

Involvement of users in Spoken Language Dialogue System design could have been limited to recording example teacher-student dialogues as a guide to grammar and lexicon development; or to asking prospective end-users what kind of functionality the system should have. In addition, we involved users at a meta-level, to decide whether spoken dialogues were appropriate (happily, teachers and learners agreed they were!), and what generic forms such dialogues should take (in our case, the exercise and feedback types). We encourage others developing SLDSs to consult meta-level domain experts from the outset, asking whether they really want a spoken language dialogue system, and what form(s) the dialogue should take. Users may indicate a preference for inclusion of other forms of interaction and processing, as well as speech input/output. This suggests a need to revisit the definition of "dialogue": other SLDS developers should also consider the merits of multimodality as an adjunct to pure spoken language dialogue, particularly given that current systems are not capable of accurately handling unconstrained English.

## References


Atwell, Eric. (1999) *The Language Machine.* London: British Council.

Atwell, Eric, Herron, Dan, Howarth, Peter, Morton, Rachel, Wick, Hartmut. (1999) Pronunciation training: requirements and solutions. ISLE Project Report D1.4, University of Leeds. http://www.nats.informatik.uni-hamburg.de/isle

Churcher, Gavin, Atwell, Eric, and Souter, Clive. (1996) Dialogues in Air Traffic Control. In Proceedings of 11th Twente Workshop on Language Technology, Netherlands.

Churcher, Gavin, Atwell, Eric, and Souter, Clive. (1997) The semantic/pragmatic annotation of an Air Traffic Control corpus for use in Speech Recognition. In Ljung, M (editor), Corpus-based Studies in English: Papers from the 17th Conference of the International Computer Archive of Modern English (ICAME), pp. 353-374, Rodopi, Amsterdam.

Civil Aviation Authority. (1992) Radiotelephony Manual (CAP 413), CAA, London.

Collins. (1994) Collins English Dictionary. (3rd edition) HarperCollins, Glasgow.

Dalby, J, Kewley-Port, D and Sillings, R. (1998) Language-specific pronunciation training using the HearSay system. In Proceedings of STILL98: Speech Technology in Language Learning, Marholmen, Sweden.

Dix, A, Finlay, J, Abound, G, Beale, R. (1998) Human Computer Interaction. (2nd edition) Prentice Hall.

Dybkjaer, L, Moller, M, Bernsen, N, Grosse, M, Olsen, M, Schiffrin, A. (2000) Annotating Communication Problems Using the MATE Workbench. In Proceedings of LREC2000: Language Resources and Evaluation Conference, Athens. See also http://www.linglink.lu/le/projects/mate

EAGLES. (1995) Evaluation of Natural Language Processing Systems. European Advisory Group on Language Engineering Standards - Final Report EAG.EWG.PR.2. http://www.issco.unige.ch/ewg95/ewg95.html

EPSRC. (1998) Devising a strategy for the support of human factors. http://www.epsrc.ac.uk/documents/programmes/highlights/itcs/strategypaper.htm

Eskenazi, M and Hansma, S. (1998) The Fluency pronunciation trainer. In Proc. StiLL98: Speech Technology in Language Learning, Marholmen, Sweden.

Herron, Dan, Menzel, Wolfgang, Atwell, Eric, Bisiani, Roberto, Daneluzzi, Fabio, Morton, Rachel, Schmidt, Juergen (1999) Automatic localization and diagnosis of pronunciation errors for second language learners of English. In Proceedings of EUROSPEECH'99.

Leech, Geoffrey. (1992) 100 million words of English: the British National Corpus (BNC). In Language Research journal, Vol.28 No.1, pp1-13.

Meekings, Alan. (1999) Discussion of Virgin Rail Telesales dialogue structure, private communication.

Menzel, W, Atwell, E, Bonaventura, P, Herron, D, Howarth, P, Morton, R, and Souter, C. (2000) The ISLE corpus of non-native spoken English. In Proceedings of LREC2000: Language Resources and Evaluation Conference, Athens.

Brett, Steven. (1999). Discussion of GE Capital Telesales dialogue structure, private communication.

Souter, Clive, Atwell, Eric, Howarth, Peter. (1999) A spoken English learners' corpus. presented at ICAME'99: International Computer Archive of Modern and medieval English, Freiburg, Germany.

Wilson, Andrew, and Rayson, Paul. (1993) Automatic content analysis of spoken discourse. In Souter, C, and Atwell, E (eds.). Corpus-based Computational Linguistics: Papers from the 12th Conference of the International Computer Archive of Modern English (ICAME), pp.215-226, Rodopi, Amsterdam.

Witt, Silke, and Young, Steve. (1997) Language learning based on non-native speech recognition. In Proceedings of EUROSPEECH97.

Wyard, Peter, and Churcher, Gavin. (2000) All channels open: multimodal human-computer interfaces. British Telecom Technology Journal, Vol.18 No.1 (Millennium Edition) $http://www.bt.com/bttj/vol18no1/today/papers/p_wyard/abstract.htm$