CUSTOMISATION AND QUALITY ASSESSMENT OF SPOKEN LANGUAGE DESCRIPTION

J. Bruce Millar,

Computer Sciences Laboratory, Research School of Information Sciences and Engineering
Australian National University

ABSTRACT

This paper describes a world-wide web based system which allows a speech data corpus developer to interact with a system for the comprehensive description of a spoken language data corpus. The interface at the browser allows the user to define the speaking environment that is to be described and then progressively to describe it and all the data files arising from it in a modular fashion. The quality of the description can be tested incrementally and the feedback generated used to further update the description. The user is effectively guided through the necessary modules in a way that reminds but does not demand adherence to a strict pattern. The complete data description can be displayed on screen and also downloaded in a simple text form to the user. Users may register as clients of the system in which case they can store descriptions on the system and upgrade these descriptions at any time.

1. INTRODUCTION

A proposal for a structure to support a comprehensive description of spoken language data has been presented previously under the mnemonic NSLD (Notional Spoken Language Description) [1]. That proposal relied on a world-wide web publication [2] for its detailed content which in turn was based on a sequence of proposals arising from the initial idea presented in 1992 [3]. The present paper addresses some of the operational challenges which impede the understanding and hence useability of such a descriptive system. Specifically this paper looks at the issue of the management of the complexity of a comprehensive description and the introduction of layers of complexity which can be used to reduce the cognitive load on the inexperienced user in particular; it looks at the issue of iterative building of the description and an associated feedback mechanism in the form of an on-line quality assessment; it also exemplifies the formal containment of the description in a database which can export its contents in forms that can be readily exploited by the user.

2. REVIEW OF CSLD

The comprehensive spoken language description (CSLD) system is distinct from the increasing array of transcription systems which have been developed to describe the sequence of acoustic-phonetic elements which comprise a spoken language utterance. CSLD can even be regarded as all the more essential because of the existence of this plethora of transcription schemes. This is because the CSLD provides the infrastructure whereby many aspects and even alternative schemes can be integrated into a single system within which the nature of each scheme, or more generally the nature of each segment of the corpus, is made explicit.

CSLD segregates the description of a speech event into three domains which can be regarded as analogous to the roots, the trunk and the branches of a descriptive tree. The roots of a speech event description are the pre-existing conditions out of which the speech event arises, including all aspects of the speaking environment such as the description of speakers, their tasks, the transducers, and the physical environment within which all these components exist. The trunk of a speech event description is the precise specification of the collection of data streams which are generated by the transducers when the event is recorded. The branches of a speech event description are the precise specification of multiple analyses and annotations using different methods and schemes which are attached to the measured data streams.

CSLD provides a structure whereby all these essential components of a comprehensive description may be linked to each other. A researcher or technologist who uses a CSLD-described corpus should be able to access all relevant data about the item of his/her interest within the corpus.

3. MANAGEMENT OF COMPLEXITY

The issue of complexity of the CSLD system has been a barrier to its acceptance as it appears to be too comprehensive for the large majority of situations. Although CSLD is essentially modular in design, it is of concern that it be presented in a form that allows the user to define the amount of complexity required

for a particular descriptive problem. A hyperlink structure commends itself for this purpose as both surface and deep, or alternative, structure can be nicely separated.

3.1. Levels within CSLD

The root, trunk, and branch levels of CSLD are explained in [1] and the low-level details supporting them are explained in [2]. The analysis in this paper relates to the interface between the general user of this scheme and the presentation of different levels of detail.

The surface level of the presentation reflects an analysis of what is essential to tell the descriptive story and to provide access points to all that lies below that level. This level arises from the need to present a cohesive logical framework that is easy to understand and which is uncluttered with unnecessary detail. At this level full explanation must be given on the screen. Lower level descriptors need to be separate and accessed via a hyperlink mechanism to allow both a greater level of descriptive power and also to access "help" facilities designed to explain what is intended at these lower levels on a "need to know" basis.

3.2. Hiding unwanted detail

Unwanted detail is a barrier to understanding but it is also a distractor in a task that may be perceived as more complex than it really is. A major factor in the successful introduction of a novel scheme such as CSLD is allowing the user the maximum control over use of the scheme in a way that does not compromise its essential character.

3.3. Formal Containment

A descriptive scheme like CSLD may be contained in a number of ways. Up until now it has been presented as constrained by an extensible <key> plus <value> formatted document. It may be that this will be superceded by an SGML-like structure which copes with linked modules of information rather than a contiguous document. For the system described in this paper we move to a relational data structure in which modules of description are implemented in relational tables.

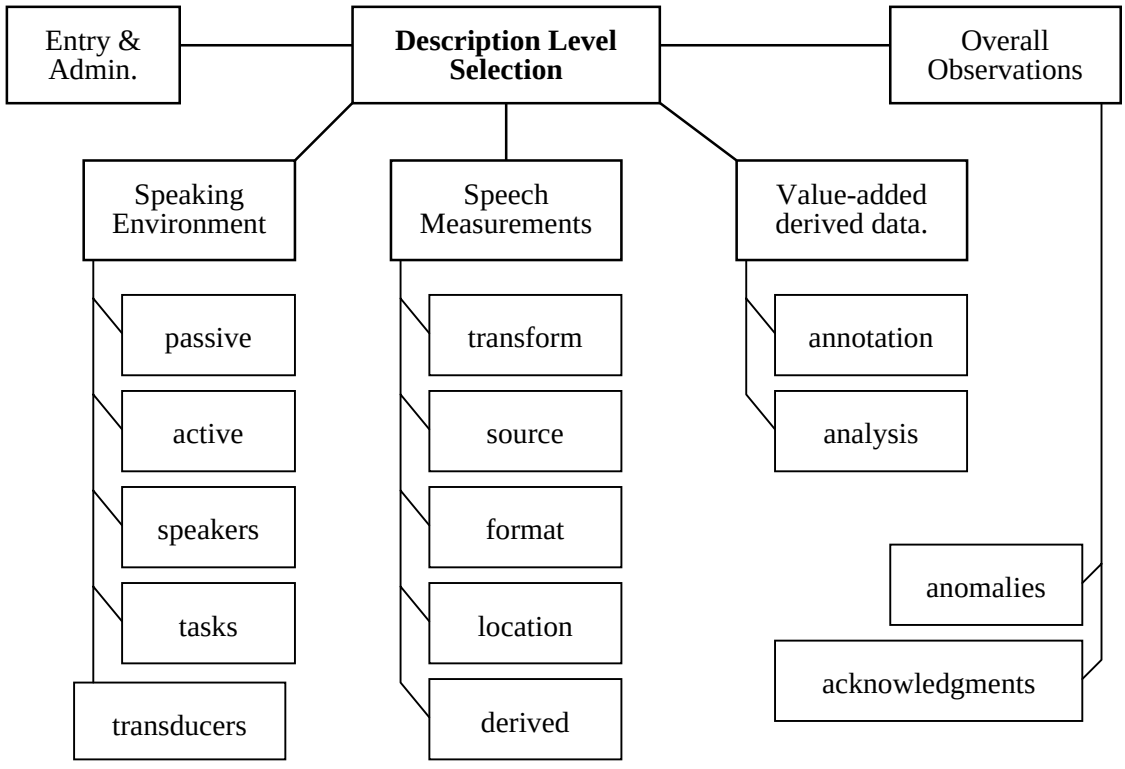

Figure 1. Modular construction of the CSLD on-line system (for explanations of the named boxes see the text).

4. ON-LINE SYSTEM

The CSLD on-line system has been implemented as a world-wide web interface to a relational database [4]. The CSLD front administrative page invites the user to "explore", "register", or "logon". "Explore" allows the user to browse the site and to understand the structure of spoken language description offered by CSLD. "Register" allows the user to be granted a relational database account on the site. "Logon" gives password-protected entry to the site to a registered client and enables such clients to store descriptive data on the site and to retrieve it at will via the WWW interface. Registered users have the option to access an existing description, create a new description or delete old descriptions. Once a description is selected as the current description each page is labelled with the client's name and the description identifier. This overall description identifier may be changed only by returning to the description selection page.

Each page is labelled with the "current settings" of the system and "help" links to remind the client how to change them. Broadly speaking, "SELECTION" pages allow the client to both navigate freely through the descriptive space and to establish the "current settings" which are then active when entering "DESCRIPTION" pages within which specific variables within the overall description may be viewed or updated.

4.1. Descriptive Level Selection

The "descriptive level selection" is the pivotal page of the system. It is entered immediately after a specific description is selected and can be directly returned to from any part of the system. It allows the selection of one of the three levels of description reviewed in section 2, namely the roots of the description or the "speaking environment" level, the trunk of the description or the "speech measurements" level, and the branches of the description or the "value-added derived data" level plus an extensible "overall observation" level. These may be entered in any order even though the logical order is suggested by the screen layout.

4.2. Speaking Environment Selection

The "speaking environment selection" introduces the client to the "roots" of the description. The client may choose in any order to initiate or update the description of "speakers", "tasks", "transducers", the "active" non-speech sources within the environment, or the overall description of the "passive" environment.

- **Speaker Selection** allows creation and deletion of speaker descriptions, the activation or suspension of these descriptions with respect to the current speaking environment, and contains links to the "passive environment" for the definition of the speaker position for the currently activated speaker description., and to "speaker description" pages for the viewing or modification of that description.
- **Task Selection** allows creation and deletion of task descriptions, the activation or suspension of these descriptions with respect to the current speaking environment, and contains links to the "passive environment" for the assignment of tasks to specific speakers within that environment, and to "task description" pages for the viewing or modification of that description.
- **Transducer Selection** allows creation and deletion of transducer descriptions, the activation or suspension of these descriptions with respect to the current speaking environment, and contains links to the "passive environment" for the definition of the transducer position for the currently activated transducer description., and to "transducer description" pages for the viewing or modification of that description.
- **Active Source Selection** allows creation and deletion of active source descriptions, the activation or suspension of these descriptions with respect to the current speaking environment, and contains links to the "passive environment" for the definition of the position of the active source for the currently activated active source description., and to "active source description" pages for the viewing or modification of that description.
- **Passive Environment Description** allows the client to describe the general characteristics of the physical space, including size, shape, temperature, humidity, auditory and visual characteristics, as well as the positions and orientations of all active elements (transducers, speakers, and other active sources) in the environment, and the assignment of tasks to speakers.

4.3. Speech Measurement Selection

The "Speech measurements selection" introduces the client to the "trunk" of the description. The client may choose in any order to initiate or update the description of the data files produced by recording from the transducers during the speech event triggered by the tasks given to the speakers. Once again five options are presented for the description of the data stream created by each of these transducers.

- **Transformation Selection and Description** allows the creation and deletion of digital transformation descriptions, the activation or suspension of these descriptions with respect to the current data stream, and their modification to reflect any post-digitisation transformations of the data.

- **Data Source Description** allows the description of the creation date, time, and place, and the size of the current data stream.
- **Data Format Selection and Description** allows the creation and deletion of data format descriptions, the activation or suspension of these descriptions with respect to the current data stream, and their modification to reflect general file format information with additional and independent sections for “binary”, “ASCII”, and “block” data formats.
- **Data Location Selection and Description** allows the creation and deletion of data location descriptions, the activation or suspension of these descriptions with respect to the current data stream, and their modification to reflect the data file name plus relevant descriptors for “published”, “on-line”, “WWW accessible”, and “physical” data corpora.
- **Derived Data File Selection and Description** allows the creation and deletion of derived data file descriptions, the activation or suspension of these descriptions with respect to the current data stream, and their modification to reflect the existence of analytic and annotation data which represent the potential for added descriptive value.

4.4. Value-added Derived Data Selection

Analysis and Annotation Selection introduces the client to the "branches" of the description. The client may choose in any order to create or modify the description of analysis or annotation files produced by algorithm or by expert examination of the data files. Two options are presented for the description of the these files.

- **Manual Annotation Selection and Description** allows creation and deletion of manual annotation descriptions, the activation or suspension of these descriptions with respect to the current annotation file, and their modification to reflect the date of annotation, the skill of the annotator, the annotation method, the annotation environment, and the annotation scheme used.
- **Analysis Selection and Description** allows creation and deletion of analysis descriptions, the activation or suspension of these descriptions with respect to the current analysis file, and their modification to reflect the details of the algorithm used in performing the analysis.

5. ON-LINE ASSESSMENT

An essential design feature of the on-line system is the ability of a client to build a description incrementally and to receive feedback on the status of his/her description. This facility is under development at the time of writing and yet promises to be a most important component of the CSLD system. It is designed to provide the following functions.

- **On-line assessment** can be accessed from any point in the descriptive process.
- **Overall complexity** of the description being attempted can be rated.
- **Balance of detail** in different parts of the description can be monitored.
- **Missing detail** according to criteria set by the client can be used to alert the client.

This on-line assessment facility will enable the client to not only monitor the progress and coverage of the current description but also will provide a means whereby an overall assessment of the descriptive quality of a corpus may be publicised for the benefit of potential users.

6. CONCLUSION

The CSLD on-line system has contributed to the concept of spoken language description in a material way. It is not simply an implementation but rather it reveals the scheme in a new way which is driven by the act of describing rather than by the end-product of a description. This approach allows the WWW client to explore the CSLD system in a modular way that suits his/her own needs avoiding excess complexity and gaining feedback on his/her progress against personally defined criteria.

7. ACKNOWLEDGMENTS

The expertise of Arthur McGuffin in the design of relational data models and their flexible access via the world-wide web is gratefully acknowledged. Without his untiring modifications of software, the evolution of this implementation of CSLD would not have been possible.

8. REFERENCES

1. Millar, J.B. A Structure for Comprehensive Spoken Language Description, 1st International Conference on Language Resources and Evaluation, Granada, Spain, 1998. Pp.1303-1308.
2. Millar, J.B. *NSLD 98 Definition Profile* URL= http://cslab.anu.edu.au/~bruce/NSLD
3. Millar, J.B. The Description of Spoken Language, Proc. 4th Australian International Conference on Speech Science and Technology, Brisbane, 1992, pp.80-85.
4. Millar, J.B. *The CSLD on-line system* URL= http://cslab.anu.edu.au/CSLD