A web-based transcription tool

M Luisa Garcia Lecumberri[1], John Maidment[2], Martin Cooke[3], Anders Ericsson[4], Mircea Giurgiu[5]

[1]University of the Basque Country, Spain

[2]University College London, UK

[3]University of Sheffield, UK

[4]University of Stockholm, Sweden

[5]Technical University of Cluj-Napoca, Romania

E-mail: fipgalem@vc.ehu.es, johnm@phonetics.ucl.ac.uk, m.cooke@dcs.shef.ac.uk, anders@ling.su.se, mircea.giurgiu@com.utcluj.ro

ABSTRACT

Transcription raises speakers' awareness of sound systems and, in the case of language learners, of pronunciation errors. It is also a valuable diagnostic technique in pronunciation competence assessment. Nevertheless, transcription requires extensive practice and feedback, making heavy demands on tutors. Autonomous transcription learning can be difficult and frustrating due to the lack of contextual variation information and personalized feedback. We describe a web-based tool (www.wtt.org.uk) for transcription practice, which may also be used for pronunciation training and diagnosis. It incorporates automatic, fast and personalized feedback, making exercises customisable and varied. The tool performs an optimal alignment of student versus model transcriptions using a dynamic programming algorithm, modified to handle alternative pronunciations. It computes a summary of errors and their locations within a student's transcription. The current version contains materials for British English and Castilian Spanish and is readily adaptable to additional languages.

1. INTRODUCTION

The idea of creating a web-based transcription tool to provide personalized and automatic feedback on students' transcription attempts originated as an answer to observed teacher and student needs, such feedback being essential to students but very time-consuming for tutors. An early prototype [1] operating standalone demonstrated the feasibility of the project. A recent web version [2] (www.wtt.org.uk) allows student self registration and monitoring by tutors at a distance, facilitates sharing of transcription resources globally, the creation of a multilingual database of transcription errors for research purposes, platform independence and easier management of the tools (e.g. release of new versions and correction of errors).

The standalone version has been in use with a large group of students at the University of the Basque Country while the first web version was used in the 2002 session. Those versions focused on English transcription. In this paper we explore the pedagogical, linguistic and technical issues raised by its extension to other languages, in particular, to Spanish.

2. PEDAGOGICAL ISSUES

In this section, we emphasize the pedagogical interest of a transcription tool for the native speaker (NS) and the foreign language learner (FLL) with particular attention to English and Spanish, the languages currently operative in WTT.

Transcription is an exercise which demands extensive practice [3], whose main virtue is the raising of phonemic and/or phonetic awareness. Thus, for the NS, it is a way of exploring and reflecting on her sound system, of making explicit her native competence [4] and of highlighting the degree of correspondence between orthography and pronunciation. This is particularly important in a language such as English due to the well known opacity of its orthography [5]. In general, transcription exposes connected speech phonemic and allophonic phenomena of which NSs are normally unaware. It is also valuable as a diagnostic technique for the tutor's assessment of pronunciation competence [6].

For the FLL, transcription reveals the target language (TL) sound system, as an alternative to purely imitative acquisition, by directing the learner to a more analytic approach to pronunciation learning. In this sense it may be argued that such an approach is not suitable for all learners (for instance, children or purely naturalistic, intuitive learners). However, we believe that WTT's customizing possibilities would also make it useful for non linguistically sophisticated learners (allowing focus on just one or two sounds at a time, for example).

In addition, FLLs share some of the benefits conveyed by transcription already highlighted for NSs. Transcription

raises awareness of the learners' interlanguage sound system and makes explicit the divergences and similarities between interlanguage and TL pronunciation. On the issue of pronunciation vs. orthography, it may be thought that since the relationship between the two is relatively predictable in a language such as Spanish, there is little point in extending WTT to this language. However, as will be seen below, the tool is flexible enough to allow for the inclusion of different degrees of allophonic detail. Thus, for Spanish FL learners, the transcription task not only aims to emphasize phonemic contrasts but also highlights allophonic and phonemic processes -such as plosive lenition or inter-word vowel liaisons- which are below the NS's level of awareness,.

3. MULTILINGUALITY

Until recently, WTT has been used exclusively for the transcription of British English (RP). However, from the outset, one of the main design goals for WTT has been support for transcription of various languages. We have recently started to investigate the issues raised by multilinguality. Our focus here is on transcription and phonetic processes in multiple languages rather than on the internationalisation of the tool itself e.g. multilingual prompts and help. Multilinguality is a desirable aim because it broadens the tool's audience and provides a more extensive test of the tool's technical design. Additionally, the use of the same tool for the transcription of several languages highlights the unifying role of the IPA conventions whilst at the same time displaying transcription differences across languages.

In this section, these questions are examined by comparing the use of WTT for English and Castilian Spanish. Some observations on the extension of WTT to Swedish and Romanian are also included. For illustration, English and Spanish transcriptions and orthography are shown in figures 1 and 2.

The clearest difference between the two is the much higher degree of transcription-orthography correspondence for Spanish, and the consequent differing aims of transcription in the two languages, as we have already noted.
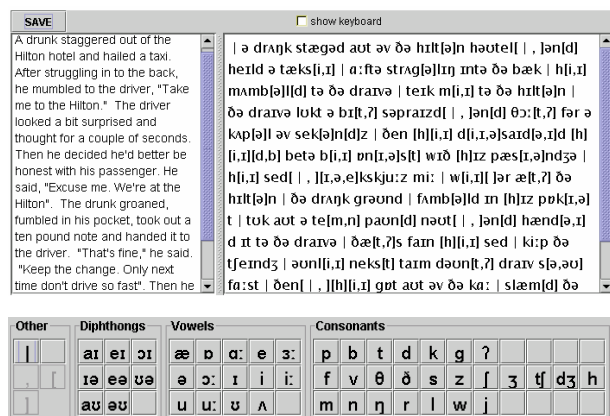

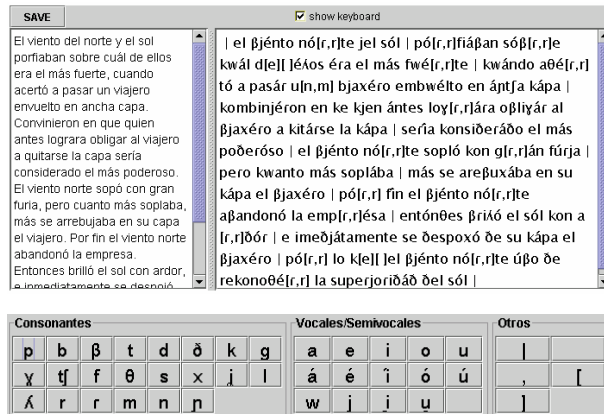**Figure 1**: English transcription sample in WTT.

Figure 2: Spanish transcription sample in WTT.

A glance at the figures shows that while the English version is a phonemic or broad transcription, the Spanish one includes a number of allophonic variants. The Spanish transcription can be considered semi-narrow, as recommended in [7] for foreign language teaching. It should be noted that the level of detail included in the transcription is a matter of choice for the linguist, within the boundaries of the tool's current technical capabilities (which are discussed in section 4). Thus, for English, only phonemes and phonemic level processes are represented. On the other hand, the Spanish transcription includes voiced plosive lenited realizations ([β, ð, ɣ]), which are more frequent than their plosive variants and therefore a possible didactic concern.

The usefulness of WTT for Swedish is strongly related to its large vowel inventory. Phonologically, Swedish has 9 distinctive vowel qualities but, in addition, it has distinctive vowel quantity and although its most important acoustic correlate is duration, vowel quality also plays a role. This means that, in principle, a speaker of Swedish must master 18 different vowel qualities to sound native. An added difficulty for FLLs is the fact that some of these vowels are comparatively rare. The Swedish sounds known to cause problems for most learners are the front rounded vowels /y, ø/ and the mid rounded vowel /ʉ/. In the UPSID database, the most common of them, /y/, only occurs in 5% of the 451 languages in the database. The other two vowels are even less frequent (/ø/ 3% and /ʉ/ 1%). It is likely that the greatest impact of WTT will be on helping raise learners' awareness of these sounds, the relationship between them and their orthographic representation and eventually help towards their discrimination.

As for Romanian, there is quite a good orthography-pronunciation correspondence (with some exceptions such as 'eu' <I> /ieu/ and not /eu/, 'idee' <idea> /ideie/ and not /idee/, 'saši' <Saxons> /saʃ/ and not /saʃi/). Therefore, as is the case for Spanish, WTT could incorporate more detail including processes that are less transparent either to the NS or to the FLL such as palatalization. Similarly, stress could be easily incorporated with additional stressed vowel symbols given that there are only seven phonemic vowels and that stress position is fixed for

nouns ('cása' <house>, 'cásele' <the houses>) and predictable for verbs, in which stress moves from the root to the ending, depending of the conjugation ('apúc' < I take> 'apucám' <I was taking>).

The transcription of connected speech processes is left to the linguist's discretion. In WTT, we treat processes as optional variants by coding them as such (e.g. "ten pound" as /te[n,m] paʊnd/). Nevertheless, the tutor could decide to exclude them altogether or to make them compulsory, dependent upon the aims of the task. The English example includes assimilations, glottaling, elisions and linking-r. In Spanish we have included nasal assimilation and some interword vocalic liaisons ("y el" /jel/). There are obviously many other processes which have been left out. For example, we have chosen not to represent gemination or allophonic assimilations in either transcription. The former raises a problem of substitutions involving multiple symbols that will be discussed in section 4 whereas the latter simply requires an extension of the keyboard to include the appropriate allophonic symbols.

WTT also handles alternative pronunciations, as can be seen in the English transcription. This is done in the same manner as phonological processes, that is, by coding the two versions between brackets (e.g. "passenger" is encoded as /pæs[ɪ,ə]ndʒə/).

An important factor that deserves attention in transcription is stress. The Spanish transcription includes word stress while the English transcription does not. Two issues, one phonological and the other technical, contribute to its absence for English.

The phonological problem relates to the unpredictability of stress in English connected speech due to deaccenting. While in Spanish most words preserve their single lexical prominence in connected speech, in English a speaker may decide to deaccent the word or, in multiple stressed words, suppress one or all of its accents. Thus, a representation of English stress in WTT would run into the question of encoding all alternative stress patterns for each particular utterance.

Technically, the introduction of stress marking into WTT is currently problematic. At present, the process used to determine the best alignment between student and model transcriptions operates on a single tier. It does not, for example, distinguish between symbols with a segmental meaning and those which mark stress or breaks. Consequently, the full adoption of stress marking awaits developments in the alignment process itself. However, it is possible to incorporate segmental stress by the introduction of a stressed segment representation on the keyboard. Spanish stress marking has been achieved in this way. While it would be possible to apply the same approach to English, the resulting keyboard would become rather crowded with the extra 19 keys to represent all accentable monophthongs and diphthongs.

A more general solution to some of the problems of stress is to adopt a multi-tiered transcription representation in which features such as stress are encoded in a different tier to the segmental one in a manner reminiscent of autosegmental phonology [8]. Technically, this would involve significant modifications to the alignment procedure to ensure that the different tiers are aligned independently while maintaining their original locational constraints.

4. WTT IMPLEMENTATION

The Web Transcription Tool (fig. 3) employs a Java servlet based distributed architecture which accesses a relational database for the storage and retrieval of model transcriptions and student attempts.
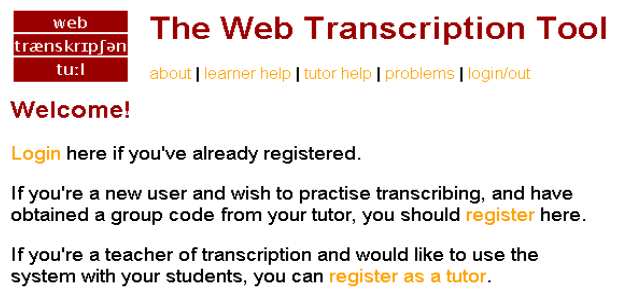

Figure 3: WTT home page.

The tool caters for both students and their tutors. Tutors are able to set up and manage logical groups of students (fig. 4, top), to which they add existing transcription tasks (fig. 4, bottom), or create new ones. Students self-register by identifying themselves with a group set up by their tutor. In this way, tutors are able to monitor the performance of their students. Students work their way through the transcription tasks, receiving different types of feedback on each submitted attempt (fig. 5, bottom).

Figure 4: Tutor view of WTT. Top: group management. Bottom: transcription management.

These transcriptions are available:

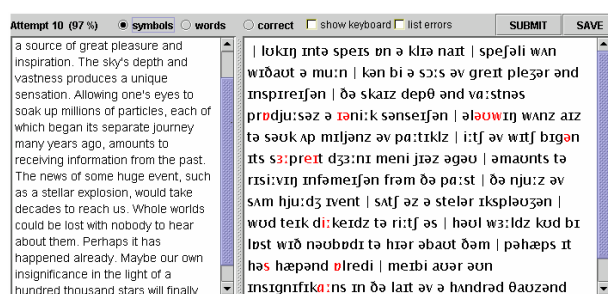| Name	Description	Attempts	Last attempt	Correct
trans1	Heavy cold	5	26-Sep-02	18
trans2	Please be patient	0		
trans3	Polishing brass	0		
trans4	Driving test	0		
trans5	Peter	0		
trans6	My mother's affairs	0		
trans7	Old age	0		
trans8	The real world	0		
trans9	Breadmaking	0		
trans10	Carnival	0		
trans11	Mushrooms	7	21-Mar-02	98
trans12	Looking into space	10	26-Sep-02	97

Figure 5: Student view of WTT. Top: list of available transcriptions. Bottom: transcription in progress, showing errors (visible in .pdf version).

At the heart of the transcription tool is a string aligner which uses a dynamic programming algorithm to determine the optimal correspondence between the student attempt and the reference transcription. The aligner computes a cost function for potential alignments by penalising symbol substitutions, insertions and deletions. Currently, the algorithm is capable of handling single-symbol substitutions and options. It cannot handle more complex alternatives or multiple representations/tiers. This limits the current applicability of the tool to unisegmental connected speech processes and lexical stress patterns where the stress mark is attached to the nucleus or to a well-defined syllable onset. There is currently nothing to prevent the aligner matching a segmental symbol to a non-segmental symbol. Future versions will correct this deficiency using a multi-tiered approach, as outlined in section 3.

Support for multilinguality within the tool is based around keyboard definition via eXtensible Markup Language (XML) [9]. XML is also used to define language-specific feedback on common transcription errors. WTT uses UNICODE [10] to define IPA symbols as well as keyboard shortcuts and example words containing each phoneme. Production of a keyboard for a new language is simply a matter of defining its symbols via their UNICODE values in an XML-format file.

Of course, there is far more to proper internationalisation of a software tool for individual languages, but our focus to date has been on the support of transcription in languages other than English rather than on, for example, localisation of prompts.

5. FUTURE WORK

Future development of the tool will focus on extending the range of languages covered, increasing the number of transcription exercises available, and algorithmic develoments to the aligner. Additionally, in order to cater for a wide range of student levels, we intend to extend the tool's range of tasks by including more basic transcriptions, such as individual sounds and words, as well as more difficult ones, such as transcription from audio stimuli (dictation). We would also like to extend feedback provision with more information on specific errors and synthesis of student transcriptions for significant audio feedback.

REFERENCES

[1] M.P. Cooke, M.L. García Lecumberri and J.A. Maidment, "A tool for automatic feedback on phonemic transcriptions", *Proceedings of VII Eurospeech,* pp. 2795-2798. Aalborg (Denmark.), 2001.

[2] M.L. García Lecumberri, M.P. Cooke and J.A. Maidment, "Transcripción Fonémica en Internet", *VIII Simposio Internacional de Comunicación Social*, vol 2, pp. 1043-1047, Santiago (Cuba), 2003.

[3] M.L. García Lecumberri and J.A. Maidment, *English Transcription Course*, London: Edward Arnold, 2000.

[4] M. Ashby, P. Ashby, J. Baldwin, F. Holmes, J. House and J.A. Maidment, "Broad transcription in phonetic training", in *Proceedings of XIII International Congress of Phonetic Sciences*, K. Elenius and P. Branderud, Eds., vol 3, pp. 170-179, 1995.

[5] M. Celce-Murcia, D. M. Brinton and J. M. Goodwin, *Teaching Pronunciation*, Cambridge: Cambridge University Press, 1996.

[6] A. Koet, "Pedagogic transcription and the acquisition of the pronunciation of English", in *New Sounds '90*, J. Leather and A. James, Eds., pp. 307-319. Amsterdam: University of Amsterdam, 1990.

[7] A. Quilis and J.A. Fernández, *Curso de Fonética y Fonología Españolas para Estudiantes Angloamericanos*. Madrid: CSIC. 1996

[8] J. Goldsmith, *Autosegmental phonology*, PhD Thesis, MIT, 1976; New York: Garland, 1979.

[9] See http://www.xml.org

[10] See http://www.unicode.org