

FluenCi and the Dynamic Speech Corpus (DSC)

Dermot F. Campbell, Marty Meinardi, Charlie Pritchard, Yi Wang

Dublin Institute of Technology (Ireland)

dermotfcampbell@gmail.com

Abstract

The EU funded FluenCi project aims to introduce learners to the importance of intonation in native-to-native communication via high-frequency collocations. Students are presented with common lexical and syntactical structures in neutral form plus several intonational variations, illustrating different speaker intentions and realised by several speech features such as stress, vowel length, pitch change and speed of delivery.

The terminology used to describe the scripted audio assets in the FluenCi database is the same as that used in DIT's Dynamic Speech Corpus, which is a corpus of native-to-native dialogues designed primarily for language learners. This corpus sensitises users to dialogue as response, rather than simply speech production, stressing the interactive strategies of speakers and making the resultant phonetic realisations accessible to learners and researchers alike.

The DSC framework is based on a unique approach to natural, informal speech and embodies this hierarchical analysis of speech in an XML database with outputs based on an innovative dialogue player where the speech and transcripts are automatically synchronised, regardless of playback speed.

The DSC can also accommodate various dialogue categories such as scripted, semi-scripted, unscripted and re-scripted, providing a full range of naturalness ranging from materials which can be used to teach vocabulary and structure to natural dialogues which capture – and make available for study – the 'messiness' of natural speech [1] in a principled fashion, including the use of a slow-down algorithm.

Whereas the natural dialogues of the DSC offer all learners an accessible model of informal native dialogue, the scripted dialogues of the FluenCi sub-programme will provide a useful scaffolding bridge between the teaching dialogues of currently available language materials and the unscripted audio assets of the DSC. Both projects are linguistically driven and prototypes are currently being developed using HTML5, which will facilitate implementation across desktop, tablet and smartphone.

1. Introduction

Speech is very different from its derivative, written language. But language spoken is not necessarily the same as spoken language. Most speech is concerned with dialogue, which is primarily about social interaction [2], yet we normally teach a reasonable vocabulary and reasonably grammatical, complete sentences. However, native speakers do not talk like that. A cursory examination of native-to-native (L1-L1) speech reveals considerable differences between carefully 'choreographed', scripted 'dialogues' and the highly dynamic and grammatically 'suspect' interactions which make up natural dialogue – Cauldwell [1] rightly talks about the 'messiness' of natural speech.

Can this messiness be taught? Probably not. But students can *learn* to cope with the dynamics of L1-L1 speech if they realise that dialogue is largely about *response* [3]. A structured approach to the seeming disorder of L1-L1 spoken interchanges is what is offered by the DIT's FLUENT Dynamic Speech Corpus (DSC) [4]. This is a collection of unscripted, L1-L1 dialogues which have been recorded, transcribed and tagged so that examples of L1 spoken language deviating from neutral

delivery can be found and studied. It also leads to a better understanding of the dynamics of the '5th Skill', i.e. speaker *interaction* in natural dialogues.

The EU-funded FluenCi project (Lifelong Learning Programme) aims to provide a stepping stone between scripted dialogues and unscripted L1-L1 exchanges. FluenCi aims to introduce learners to the importance of intonation, using high-frequency collocations. Students are presented with common lexical and syntactical structures in neutral form plus several intonational variations, illustrating different speaker intentions and realised by the Speech Features: stress, vowel length, pitch change and speed. Having been sensitised to the critical role of intonation and prosody in L1-L1 speech – according to Mehrabian [5], responsible for almost 40% of the message – students will be in a strong position to study the dynamics of natural L1-L1 dialogue and understand what L1 fluency really means.

2. Hierarchy of speech

DIT's DSC divides dialogue into 3 levels: Turns (T), Flow Sequences (FS) and Flow Units (FU), corresponding to speaker interaction, semantic communication and phonetic delivery.

[inhales]	FUA22	FSA9	TA5
yeah	FUA23	FSA10	
[inhales]	FUA24		
yeah you know	FUA25		
she did	FUA26		
she did like it	FUA27		
but there were sometimes she	FUA28	FSA11	
n- now that we have her in this other place	FUA29	FSA12	
and we are looking back at it	FUA30	FSA13	
it seems like she was	FUA31		
maybe	FUA32		
mmm	FUA33		
you know	FUA34		
[inhales]	FUA35	FSA14	
not as happy there as she might have been	FUA36		
yeah yeah sh-	FUA37	FSA15	TA6

as she could have been	FUB42	FSB12	TB4
--------------------------------------	-------	-------	-----

Do you know wh-	FUB43	FSB13	TB5
what lead	FUB44		
to that	FUB45		
do you think	FUB46	FSB14	

Fig.1. Hierarchy of speech levels

At the **Turn** level speakers demonstrate their interactive skills. Having started a dialogue, speakers are expected to respond to their interlocutor and stay 'on message'. Anything else is deemed rude or bizarre. They can politely take their turn or interrupt. Another possibility is to offer encouraging 'back-channelling' contributions to the conversation, which encourage the speaker to continue as before, or to modify their speech production in the light of the feedback provided. Finally, speakers might choose to defend their turn by trying to exclude the interlocutor. This cut-and-thrust of real L1-L1 speech is almost impossible to stage in a classroom and can only effectively be studied with the use of technology such as DIT's DSC.

The second level of dialogic speech (Flow Sequences) is that of the *semantic level*, where the speaker seeks to convey a message. Instead of producing neat sentences, clauses and sub-clauses, native speakers constantly modify their speech production in the light of feedback they get from their interlocutor in the form of back-channelling and visual clues (calculated by Mehrabian [5] to constitute 55% of the message). This is because in dialogue there is only one speaker, but *two* listeners.

The third level of natural speech – as implemented in DIT's DSC – is at the *phonetic level* of Flow Units. This is where Cauldwell's 'messiness' is most clearly evident. Because of the highly interactive nature of natural dialogue, there is rarely time to produce perfectly formed sentences. Speakers have to formulate an idea, find the words to express it adequately, marshal these words into an acceptable grammatical structure and deliver them in a manner appropriate to the unique circumstances in which they are to be uttered. All the while speakers have to defend their own turn, be mindful of what has already been said in the course of the dialogue, and achieve their social and personal objectives in the dialogue. Instead of pristine structure we get false starts, repetitions, humming-and-hawing while the right word is sought, changes of mind, structure repair etc., i.e. Cauldwell's 'messiness'. This level of language poses little difficulty for the native speaker, but acts as a formidable barrier to the poor learner!

3. Formulaic sequences

Not only does the actual choice of word count for only 7% of the conveyed message, but almost 60% of what we say in real dialogues is retrieved from memory rather than constructed according to grammatical and syntactical rules [6]. We frequently use off-the-peg collocations and idioms because it is easier, and because what is important is not so much *what* we say, but *how* we say it. Humour and innuendo often depend on intonation, and questions can morph into orders, depending on how they are spoken.

Given the statistical and communicative importance of high-frequency collocations and phrases, the FluenCi project uses them to introduce learners of English to the role of intonation and prosody in L1-L1 speech. 200 of the most frequently naturally occurring collocations have been chosen to form a PHRASECON – a series of phrases without which speaker interactions seem dry, unnatural and lacking in fluency.

4. FluenCi - positioned between scripted and unscripted dialogues

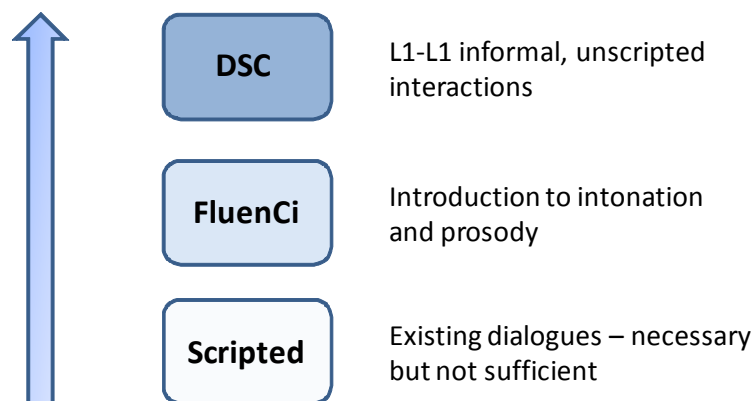


Fig.2. FluenCi between scripted and unscripted dialogues

Classroom dialogues are absolutely necessary, but they in no way prepare learners for real speech in an L1 English speaking community.

The main aim of FluenCi is to bridge the gap between scripted and unscripted dialogues, i.e. between *teaching* dialogues and *natural* dialogues. By using high-frequency phrases as a basis for introducing learners to the important role of intonation and prosody in L1-L1 interchanges, FluenCi enables students to progress beyond the 7% of communication which is expressed in the choice of word and enter the world of the 38% where the *manner* of speaking is more important. Since the tags used to describe communicative features highlighted in FluenCi are derived from the DSC, use of the FluenCi

programme will facilitate learners in coping with wholly natural, unscripted dialogues as exemplified in the FLUENT DSC.

There is another, unique form of dialogue called 'Second Pass' (or 're-scripted') which has been developed by DIT. This is a version of the unscripted dialogue which has been redrafted to maintain the natural interchanges of natural dialogues, but based on Flow Sequences (= semantic level) rather than on Flow Units (= phonetic/prosodic level). Second Pass is perhaps a more appropriate model for users of English in the international sphere - rather than in an English speaking community - or for use by linguistically less gifted students, since its syntax has been 'tidied up'. It could therefore be a useful L2 production model for users of English as an International Language.

5. From theory to practice

The academic partners in FluenCi are DIT and UNED, the Universidad Nacional de Educación a Distancia – the Spanish equivalent of the UK's Open University and the largest university in Europe. UNED's Structured Learning Materials are designed for distance learning and offer the innovative FluenCi approach to speech and dialogues in a conventional fashion, with book and attached audio CD. This current paper, however, deals with DIT's online delivery of its Structured Learning Materials using HTML5 and CSS3 on the internet.

Each Unit, or Phrase Module, revolves around one of the high-frequency phrases also found in UNED's SLMs and deals with the phrase used in a neutral context or in 2 or 3 variations in which speaker attitude or the discourse function of the phrase is made clear. These 3-4 versions of the phrase are drawn from MiniLogues (i.e. short, contextualising dialogues) which outline the pragmatic context for the neutral and marked versions of the phrase being studied. The associated audio recordings (available at normal speed and also slowed down – without tonal distortion – for more efficient study of intonational patterns) are accompanied by a series of 6 student activities as outlined below.

These Activities are chosen so as to allow a progressive development of students' awareness of the role of intonation in L1-L1 interactions, what they signify, and how this is achieved, using the following schema:

1. *Tuning in* to basic aspects of the speakers, such as their variety of English.
2. *Noticing* basic aspects of the neutral and marked Phrase Sequences such as tone, positivity/negativity and 'energy'.
3. Identification of the *Speaker Intention* (user-friendly descriptions are provided).
4. Discovery of the *Speech Features* used to convey the marked versions (stress, intonation, speed or vowel lengthening).
5. *Pragmatics*: students find that the reason each version was spoken in that particular manner derives from the context in which it was spoken.
6. The main idea of FluenCi is to demonstrate how native dialogic fluency is achieved by the appropriate use of the Speech Features exemplified in the exercises. This Activity heightens awareness of the *Linguistic aspects of the Speech Features* studied.
7. Having been made aware of the subtleties of the suprasegmental level of speech, students are now invited to *demonstrate* what they have learned in the other exercises.

In order to bring order to this extensive set of exercises, an intuitive interface is being designed to guide the student through the linguistic activities. All Units are dealt with in the same manner, so the learning curve for the user will be shallow. Since the interface is designed for the world-wide web, it could also be adapted for delivery to other platforms such as tablets or smart mobile devices with internet access, the essential difference being merely screen size.

A player is currently being developed which will allow users to play the MiniLogues (and the neutral and marked Phrase Sequences derived from them) at normal speed and slowed-down speeds of

80%, 60% and 40%. This allows students of different linguistic ability to find a speed compatible with their linguistic level and to highlight the intonational patterns involved in the marked versions of the Phrase Sequences. The use of slow-down is analogous to using fast photography to slow down a tennis serve or a golf swing, so as to better study the technique involved. Playing back the 'audio slow-motion' versions and repeating them, will allow learners to imitate the native intonation patterns of the normal speed versions, since there is no tonal distortion involved in the slowed samples. Also, slowing down speech which has been spoken at normal native speeds allows the melody of the speech to be highlighted – and this is precisely where 38% of the communication value is encapsulated.

6. Technical considerations

Given the complexity of the linguistic Activities to be implemented, HTML5 is the chosen development platform, as this will support the development of a customised player rather than external plug-ins such as Flash, and the new presentation functions supported will allow more visually appealing designs to be implemented. There are of course differences in the level of support for HTML5 in the major browsers, so care has to be taken to ensure that this is covered in the final implementation.

By 2013 it is forecast there will be more mobile web connections than desk-top connections so it is important to be able to implement the solution across mobile platforms in either web or native applications. New toolsets and frameworks will make this conversion task much easier without the need to learn the programming environments of the various mobile operating systems.

References

- [1] Cauldwell, R. (2002) *Phonology for listening: relishing the messy*. [Online] Available at: http://www.speechinaction.net/SPARC_Richards_publications.htm
- [2] Campbell, D. Wang, Y. Meinardi, M. Richardson, B. McDonnell, C. & Pritchard, C. (2010) Dialogic fluency — why it matters. *Proceedings of the TISLID 2010 Conference*.
- [3] McCarthy M. J. (2010) Spoken fluency revisited. *English Profile Journal. Inaugural issue*. [Online] Available at: <http://journals.cambridge.org/action/displayJournal?jid=EPJ>.
- [4] Campbell, D. McDonnell, C. Meinardi, M. Pritchard, C. Richardson, B. & Wang, Y. (2009) DIT's Dynamic Speech Corpus. *Speak Out!*, 41.
- [5] Mehrabian, A. (1981) *Silent Messages: implicit communication of emotions and attitudes*. 2nd ed. Belmont, CA: Wadsworth.
- [6] Erman, B. & Warren, B. (2000) The idiom principle and the open-choice principle. *Text*, 20.



Education and Culture DG

Lifelong Learning Programme

This project has been funded with support from the European Commission.

This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.