Stimulus design and transcription issues in cross-varietal investigation of spoken Arabic prosody.

The Intonational Variation in Arabic (IVAr, 2012-2014) project will provide an in-depth survey of the range and scope of typological variation in the intonational properties of spoken dialects of Arabic. Parallel speech data will be collected for input to phonological analysis within the Autosegmental-Metrical (AM) framework (Ladd 2008). The resulting IVAr Corpus and accompanying prosodic annotations will be made available as an online resource. This paper sets out the particular issues faces by cross-varietal prosodic investigation in Arabic, and our proposed responses to them.

Prior literature on the intonation of spoken Arabic dialects is limited, but in recent years a growing number of descriptions of individual dialects have appeared: for some varieties there is an AM analysis (Chahal & Hellmuth 2012) but for others we have only a preliminary description (Kulk *et al.* 2003, Blodgett *et al.* 2007), a description in a different framework (Benkirane 1998) or no description at all (e.g. Iraqi). Chahal (2006) effectively summarises these diverse descriptions. Various experimental investigations have looked at suprasegmentals in subsets of dialects (Biadsy & Hirschberg 2009, Ghazali *et al.* 2002, Ghazali *et al.* 2007, Yeou 2004, Yeou *et al.* 2007, Barkat *et al.* 1999) but in general the data are not in the public domain. A set of parallel data and descriptions is much needed, to facilitate meaningful comparison and as input to growing interest in Arabic dialectal variation.

IVAr aims to provide not only a first parallel dataset across nine dialects, but also i) a stimulus set to allow expansion of the corpus to additional dialects and ii) an Ar-ToBI annotation system. In this paper we thus present a brief overview of the rationale of the project, together with i) results of pilot testing of the corpus stimuli set and ii) our prototype Ar-ToBI system.

The experience of large-scale cross-varietal investigations of prosody, such as IViE (Grabe 2004) and the Atlas of Spanish Intonation (Prieto & Roseano 2010), has demonstrated that directly parallel speech data, containing utterances of the same sentence/paragraph by a number of different speakers, can be instrumental in identifying in what ways and in what contexts the intonational patterns used by speakers of different dialects differ, while also highlighting shared features. Variation in lexis, syntax and in segmental and metrical phonology across Arabic dialects means that creation of utterances which will be directly parallel in different dialects is not a simple task. It is not a solution to employ Modern Standard Arabic (MSA) terms and structures since this risks elicitation of a 'non-local' register of speech. In this paper we will present the results of ongoing pilot testing of our proposed stimulus set, which contains lexical items and structures which are sufficiently parallel across dialects to permit comparison, but which will also allow us to collect speech materials in which expected cross-varietal distinctions are present.

An important contribution of IViE was to propose a labelling tier (Grabe 2001) for fine-grained transcription of the type of cross-varietal variation relevant for English (local f0 contour), and work on French has argued for a further tier (global f0 contour) to capture variation in French, which marks only the edges of prosodic constituents (Post & Delais-Roussarie 2006). There is good reason to suspect that Arabic dialects stand at different ends of a number of prosodic continua (Jun 2005), such as density of accent distribution/pitch accent inventory size (e.g. Egyptian vs. Lebanese, Chahal & Hellmuth to appear), and marking of heads + edges, or edges only, of prosodic constituents (e.g. Egyptian vs. Moroccan, Phillips-Bourass 2012, Hellmuth to appear). Prior comparative work suggests however that although both local and global tiers may be relevant for Arabic varieties, manual annotation of both in a full corpus is impractical (Hellmuth & Chahal 2009). In this paper we will present the results of pilot implementation of the following protocol: a) the "alternatives tier" (Brugos *et al.* 2008) in an initial 'broad' transcription by two transcribers is used to identify which category labels are prone to ambiguity and in which contexts; b) the subset of data thus identified can then usefully be submitted to a narrow transcription, using IViE-style local and global f0 tiers, to finalise the particular properties of a given label in a particular variety.

References

Barkat, Melissa, John Ohala, & Francois Pellegrino. 1999. Prosody as a distinctive feature for the discrimination of Arabic dialects. *Eurospeech 99.* 395-398.

Benkirane, T. 1998. Intonation in Western Arabic (Morocco). In Hirst, D & A. Di Cristo (eds.) *Intonation systems: a survey of twenty languages.* 345-359. Cambridge, MA: CUP.

Biadsy, Fadi. & Julia Hirschberg. 2009. Using Prosody and Phonotactics in Arabic Dialect Identification. *Proceedings of Interspeech 2009 (Brighton, UK).* 208-211.

Blodgett, A., J Owens, & T. Rockwood. 2007. An initial account of the intonation of Emirati Arabic. *Proceedings of the 16th ICPhS, Saarbruecken, Germany.* 1137-1140.

Brugos, Alejna, Nanette Veilleux, Mara Breen, & Stefanie Shattuck-Hufnagel. 2008. The alternatives (Alt) tier for ToBI: advantages of capturing prosodic ambiguity. *Speech Prosody 2008 (Campinas, Brazil).* 273-276.

Chahal, Dana & Sam Hellmuth. to appear. Comparing the Intonational Phonology of Lebanese and Egyptian Arabic. *Prosodic Typology. Volume 2.* Oxford: Oxford University Press.

Ghazali, S., R. Hamdi, & M Barkat. 2002. Speech rhythm variation in Arabic dialects. Bel, Bernard and Marlin, I. 127-132. Aix-en-Provence: Laboratoire Parole et Langage. Proceedings of the Speech Prosody 2002 Conference, 11-13 April 2002.

Ghazali, S., R. Hamidi, & K. Knis. 2007. Intonation and rhythmic patterns across the Arabic dialects continuum. In Benmamoun, Elabbas (ed.) *Perspectives on Arabic Linguistics: Papers from the annual symposium on Arabic Linguistics Volume XIX: Urbana, Illinois April 2005.* 97-122. Amsterdam: John Benjamins.

Grabe, Esther. 2001. The IViE Labelling Guide. (Version 3.0) http://www.phon.ox.ac.uk/IViE/guide.html.

Grabe, Esther. 2004. Intonational variation in urban dialects of English spoken in the British Isles. In Gilles, P. & Jörg Peters (eds.) *Regional Variation in Intonation.* 9-31. Tuebingen: Niemeyer.

Hellmuth, Sam & Dana Chahal. 2009. Intonational Variation in Arabic: a transcription workshop. (Presentation at IVA09 Workshop, University of York, September 2009).

Hellmuth, Sam. to appear. Phonology. In Owens, Jonathan (ed.) *The Oxford Handbook of Arabic Linguistics*.

Jun, Sun-Ah. 2005. Prosodic typology. In Jun, Sun-Ah (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing.* 430-458. Oxford: OUP.

Kulk, F., C. Odé, & M. Woidich. 2003. The intonation of colloquial Damascene Arabic: a pilot study. *IFA Proceedings.* 25 15-20.

Ladd, D. R. 2008. *Intonational Phonology*. Cambridge: Cambridge University Press.

Llamas, Carmen. 2007. A new methodology: data elicitation for regional and social language variation studies. *York Papers in Linguistics Series 2.* 8 138-163.

Phillips-Bourass, Sara. 2012. Prosodic Focus Marking in Moroccan Arabic. Paper presented at the 26th Annual Arabic Linguistics Symposium. 1st -3rd March 2012, Columbia University, New York City.

Post, Brechtje & Elisabeth Delais-Roussarie. 2006. Transcribing intonational variation at different levels of analysis. In Hoffmann, R. & H. Mixdorff (eds.) *Proceedings of Speech Prosody 2006.* Dresden: TUD Press Verlag der Wissenschaften GmbH.

Prieto, Pilar & Paolo Roseano. 2010. *Transcription of Intonation of the Spanish Language*. Muenchen: LINCOM.

Yeou, M. 2004. Effects of focus, position and syllable structure on F0 alignment patterns in Arabic. JEP-TALN 2004.

Yeou, Mohammed, Mohamed Embarki, Salal AlMaqtari, & Christian Dodane. 2007. F0 alignment patterns in Arabic dialects. *Proceedings of the 16th ICPhS, Saarbruecken, Germany.* 1493-1496.