Book Review

Prosodies.
With Special reference to Iberian Languages

CARLOS GUSSENHOVEN
Radboud University Nijmegen
Queen Mary, University of London
c.gussenhoven@let.ru.nl

Prosodies. With Special Reference to Iberian Languages. Edited by Sónia Frota, Marina Vigário & Maria João Freitas (2005). Berlin & New York: Mouton de Gruyter.

The volume under review contains the written versions of twelve of the presentations at the first Phonetics and Phonology in Iberia (PaPI) conference. I wrote this review shortly after attending the third PaPI in Braga in 2007. Both the volume and the fact that the conference of 2003 made it into a series demonstrate that the field is thriving in Iberian studies. The PaPI conferences are firmly rooted in the Laboratory Phonology movement and exude a healthy desire to treat phonology as a conventional branch of science.

With three exceptions, the issues that are raised in the twelve chapters are approached through experiments. An eminently 'Iberian' paper is the contribution by Mariapaola D'Imperio, Gorka Elordieta, Sónia Frota, Pilar Prieto and Marina Vigário, who investigated the factors that influence the presence of a prosodic break after the subject and/or the verb in a read corpus of SVO sentences in which the branchingness of subject and object was orthogonally varied with the length of these constituents measured as the number of syllables. The investigation not only produced very clear results for the contribution of these variables, but strikingly demonstrates that these results are very

***J*ournal of *P*ortuguese *L*inguistics, 6-2 (2007), 83-91** ISSN 1645-4537

different for the languages under investigation, Catalan, Italian, Northern Portuguese, Standard Portuguese and Spanish. For instance, in Standard Portuguese length has a clear influence on phrasing, while in Northern Portuguese branchingness is a more powerful factor. The corpus that was used is not necessarily representative of the languages, and the number of speakers per language, two, should make us careful about drawing conclusions for the languages investigated, but the results strongly suggest that prosodic systems of closely related languages differ in significant ways, and that we ought to be thinking about standardized typological features for intonation, like length of phrases, frequency of accented syllables, number of different pitch accents in the language.

Lisa Selkirk's opening contribution represents another of her elegant accounts of the prosodic structure of English. It is important for the statement of the correspondences between prosodic constituents and syntactic constituents, and more than in earlier accounts she stresses the role of syntax in determining the prosodic structure. The edge-based account of focus ('Align Focus with the right edge of a prosodic constituent'), empirically dubious for English (Gussenhoven 1996), is abandoned in the face of obvious counterexamples and replaced with an account that accords sentence-level prominence to focus constituents, or FOCUS constituents, because the focus must be of a contrastive kind, as indicated by the capitals. This means that broad focus has a phonologically lower level of stress or prominence than contrastive FOCUS, a position that re-opens the issue of two discrete levels of pitch height, high and super-high, which was laid to rest by Ladd & Morton 1996. Of course, high and super-high need not correspond to the distinction between Major--Phrase level and IP-level stress, but the fact that no discreteness was to be found in the pitch domain should make us wary of thinking that discreteness can be found in a more general stress or prominence domain above the word level which is independent of accentuation. Similar wariness should be exercised when evaluating Selkirk's claim that the focus-to-prominence constraint is universal. The association between focus and general prominence is widespread in languages, but interestingly there are counterexamples, one of them presented by Hellmuth in the same volume. There are more empirical issues in this paper, such as the claim that there are equivalent boundaries after *Romans* in *The Romans, who arrived early, found a land of wooded hills* and after *garden* in *Cindy isn't planting a garden because she loves tomatoes (She has an entirely different reason for not planting a garden)*. To me, there is an intonational phrase break after *Romans* and none after *garden*.

Applying the methodology used in Swerts, Krahmer & Avesani (2002), Sam Hellmuth reports on an investigation of the realization of given constituents in Cairene Arabic, and finds that, as in Italian, contrastive adjectives or nouns are not deaccented in Cairene Arabic. The data are collected from speakers who were asked to identify objects that only vary in shape (square, circle, etc.) and colour. By manipulating the order of the objects to be identi-

fied, either the colour or the shape may be the same as that of an immediately preceding object, or both may be different. While in Dutch, speakers reliably deaccent the word containing the repeated information, speakers of Italian do not. Hellmuth's investigation is interesting for a number of reasons. First, the lack of deaccenting in constituents within NPs is extended to syntactic phrases. Swerts et al. did not investigate their subjects' speech in situations in which entire syntactic phrases contained contrastive information, but on the basis of Ladd (1996) they should have been expected to deaccent given phrases. To investigate this issue, Hellmuth developed the shapes-and-colours methodology further by designing a murder mystery game in which subjects are made to form sentences specifying the victim, the murder weapon and the place of the murder, with order of presentation suitably manipulated so as to cause one of the constituents to contain contrastive information. Speakers of Cairene Arabic appeared not to deaccent even at the sentence level. Second, following Swerts et al., Hellmuth ran a perception test with a context retrieval task, to find that listeners did not successfully identify preceding contexts, except one, who performed above chance, and who – intriguingly – was an advanced L2-English speaker. This is the third instance I have come across of foreign listeners outperforming native listeners, the other two being reported in Kubozono 2002 and Cutler 2007. I will refrain from attempts to compare the cases, but one interest here is in the way speakers and listeners negotiate what is and is not reliable information in the signal: not everything is used in speech perception. Thirdly, in a debriefing session, Hellmuth checked to see if corrective focus leads to deaccenting by eliciting sentences like *It isn't X, but Y*, only to find that speakers clefted the Y, then negated the X. Her speakers rejected in situ corrective focus. It would seem then that speakers of Cairene Arabic have no grammatical ways of deaccenting a word and that marking new and given information is not what they concern themselves with. This is typologically important information in the sense that there doesn't seem to be any competition from a lexical tone contrast or from other prosodic contrasts that might be at risk if deaccenting was used. In other words, the behaviour cannot be explained in terms of contrast preservation (cf. Gussenhoven 2004: 73).

Teresa Cabré and Pilar Prieto investigate the conditions under which V-V sequences at word boundaries in Catalan are reduced, and conclude that it is nuclear stress on the word-initial vowel that works against the reduction of the preceding vowel: *aixì obre* '(S)he opens it' is pronounced as a disyllabic [i.ɔ], but *aixì obre porta* '(S)he opens the door' is [jɔ], because the first syllable of *obre* no longer has the nuclear pitch accent. Other factors determine whether the reduction takes the form of glide formation or vowel deletion, but proximity of stresses is not a factor, as the authors observe, puzzled as to why other accounts claimed that it was. There is an elegant OT analysis of the facts, in which the crisp alignment of the nuclear-accented word with a foot prevents the first syllable from acquiring an onset from a preceding word. There is a challenge to their solution in the paper by Miquel Simonet, which

shows that *within words* the preference for non-reduction is greatest in the main stressed syllable. Simonet, who only investigates word-internal hiatus in Catalan and Spanish, explains the results of his questionnaire on the basis of available duration: the more duration, the less reduction. If the first V of the hiatus is three syllables away from the word stress, as in *dialogué* 'I dialogued', reduction is most likely; if two, as in *dialógo* 'I dialogue' it is less likely, and if one, as in d*i*[*á*]*logo* 'dialogue', it is least likely. To the extent that these preferences reflect actual speech behaviour, they can be explained in the spirit of Cabré and Prieto by assuming crisp alignment of feet (to get [di.álogo]) as well as a preference for disyllabic trochees (to get [di.a.lógo] and [dja.lo.gé]). But Simonet is likely to object to this approach, as he maintains that the contrast between the monosyllabic and disyllabic pronunciation isn't just variable, but rather not obviously discrete, with the grammar taking a liberal view of the difference.

Native speaker judgements as collected by Simonet should be handled with care: they represent the reflections of speakers on natural phenomena they are not supposed to know much about. One might think that Adelina Castelo's contribution ('The perception of word primary stress in Portuguese') is about the perception of word stress in Portuguese. It isn't. It is about lay listeners' judgements about where the primary word stress is in auditorily presented words. Strictly speaking, the results have as little to say about the processing of stress, the location of stress, or the phonology of stress in Portuguese as an investigation into native speakers' judgements of the segmental structure of words has to say about the segmental structure of words, or indeed as much as a historical survey of linguists' analyses of the structural location of conjunctions has to say about the structural location of conjunctions. A good linguistic theory will be able explain the data from such surveys, but their status of evidence is less clear. Humans are aware of the structure of their language only to a very shallow extent, even if that awareness varies from structural element to structural element. (For many languages, consonants are more accessible than tones, and syllables are more accessible than feet.) Having said that, Castelo's investigation produces results that many linguists will find a joy to explain. Thus, people find it somewhat more difficult to identify regular stress than to identify stress as determined by suffixes that are lexically marked for stress. The point I am trying to make is that this fact requires an explanation, which is no doubt that marked stress is more salient than regular stress, which explanation in its turn strengthens the theory that proposes that some stresses are regular and others morphologically induced. It does not show, for instance, 'the role of [...] morphological cues [...] in the perception of stress' [p 161], only the role of such 'cues' in the formation of people's opinions about where a primary stress is, whatever that notion means to them. To be honest, Castelo formulates her conclusions carefully, and I don't wish to suggest that she actually takes her results into the realm of speech processing, but I thought it was good never-

theless to identify the line which she manages to stay behind without making her position all too explicit. Let me end with an attempt at an explanation of the one finding she leaves undiscussed. In shorter words, people misidentify the word stress more frequently than in longer words, despite the greater number of opportunities which long words offer to get the stress wrong. The original motivation for including longer words was to see if people are inclined to impose a rhythmic structure over the pre-stress syllables, and misidentify as stressed that ones that come out as relatively strong. She finds no such effect, only the effect noted above. I would guess the explanation is that in shorter words, the stress stands out less than in longer words, just as it is easier to identify a tennis ball among a collection of white billiard balls more easily that if it was paired with just one billiard ball.

The question how we recognize words divides into three subquestions. First, what element in the speech signal is employed for a search through the lexicon? Is it the feature, the segment, the half-syllable, the syllable, the foot, the word, or a linguistically unanalyzed portion of acoustic information? Second, what does the search space look like? Is the lexicon a collection of banks of acoustic exemplars, of fully specified phonological representations, or of underspecified phonological representations? Third, how is a search launched? The answers to questions 1 and 2 in particular are interdependent in that it makes no sense to take acoustic information to a lexicon specified in segments, or vice versa. My money would be on (1) features (2) a featurally underspecified set of representations, with (3) being language-dependent.The language-dependence of word-recognition has been shown in the work by Anne Cutler and colleagues, who also suggested that in English searches are started at the beginnings of words that begin with a stressed syllable. The answers to the first two questions, which I have chosen because they constitute the most economical search procedure, define the FUL model by Aditi Lahiri and Henning Reetz, in which the interesting situation may arise that a feature that is detected in the signal is taken to a lexicon in which the word that is to be recognized does not have it in its specification. I have given the above sketch of what I understand word recognition is about in order to be able to place the research question that is tackled by John Kingston in perspective. The issue for him concerns the point at which the initial auditory processing of the incoming signal is influenced by the listener's knowledge of the language that the word to be recognized belongs to. The question is relevant because different models make different claims: some assume a form of language-independent processing, others assume that perception is guided by statistical knowledge of the lexicon right from the start. The language--independent processing is specified as the recognition of a 'category', but because there is no definition of what that category is, it is not easy to evaluate the research method that is applied. I took it to refer to an IPA segment, which is language-independent in the sense of independent of any specific language, though part of the world of linguistic sound. The research question,

however, suggests it is something more abstract. The test is in whether cooperative acoustic configurations, which make for greater discriminability than non-cooperative acoustic configurations in speech sounds, also make for greater discriminability in non-speech sounds. If they do, then the position for language-independent processing will have been demonstrated. Concretely, voicing of a stop closure and F1 lowering of the adjacent vowels both cause an increase in low-frequency energy, and sounds that have these properties should be highly discriminable from sounds that lack both, i.e. have voiceless closures without F1 lowering. However, sounds with the crossed configurations (voiceless closure with F1 lowering or voiced closure without F1 lowering) should be poorly discriminable. All this is true, as shown in Kingston's experiments. The test now comes in whether low-passed filtered versions of these stimuli, which no longer sound like speech sounds, reveal the same effects. Even though the answer appears to be affirmative, Kingston is very wary about the issue, and demonstrates great sensitivity to possible alternative interpretations as well as to the results produced by other researchers. His contribution not only contains a valuable survey of research into language--specific effects in perception, but an exemplary line of reasoning that leads to experiment after experiment. As will be obvious from the above, I don't follow the word recognition literature, and whether or not as a result, I wondered at the end of the paper whether it had selected the right grail. There may be more joy in attempts to determine the point at which language-specific knowledge starts to interfere with acoustic processing. By way of thought experiments, we might create two conditions, one in which subjects are told they will have to detect the difference or sameness of two noises produced by steam escaping from some engine, and one in which they are told the noises are the beginnings of words in their language. If stimuli are gated, such that subjects hear portions from the beginning of the signal but with varying lengths, the results will show more sameness in the language condition than in the steam condition at some point. If the language task is further divided into a 'different token' and a 'different type' condition, the linguistic effect could be further divided into a phonetic hearing mode and a phonological hearing mode. In cross-linguistic experiments, the question whether listeners can hear non-phonological differences may be answered, but also the question how soon listeners resort to knowledge about their lexicon. Of course, different speech sounds will require different non-linguistic rival interpretations.

Andries Coetzee reports an experiment in which he intends to demonstrate the effect of grammatical knowledge on phoneme recognition, independently of phoneme-to-phoneme transitional probabilities. He pits the recognition of the final consonants in (non-)words like *spop* and *skock* against rival consonants that do not violate the OCP, by cross-splicing the first halves of these non-words with an eight-step acoustic continuum running from *op* to *ock*. Listeners appear to be biased against hearing a repetition of the prevocalic consonant, i.e. against [p] in the *spop – spock* condition and against [k] in the

skock-skop condition. Coetzee's point here is that the bias is due to a grammatical constraint, the OCP, and not to statistical knowledge of phoneme-to--phoneme transitions. In the experimental design and in the discussion he addresses two formidable threats to this interpretation. The first is that words like *spop* and *skock* do not exist, and that the results could therefore be due to knowledge of the absence of *skVk* and *spVp* in the lexicon. To this end, he included *stVt* in the experiment, pitting it against both *stVp* and *stVk*, creating two more continua, one from *Vp* to *Vt* and one from *Vk* to *Vt*. If the same bias can be shown to exist against [t] in *stVt*, then, in view of the existence of *state, stoat, astute, stat, stout, stud*, etc., this would be an effect of the OCP which cannot be traced to statistical facts. In each of the three continua, a bias is predicted towards the consonant that is different from the pre-vocalic consonant, which in fact turns out to be true. For each continuum, the cross-over points along the three eight-point continua between the two conditions differ significantly in the right direction: [p]-[k] by 1.4, [k]-[t] by 0.29, and [p]-[t] by 0.75 scale points. However, while the results for all three continua are significant on the basis of t-tests on the responses, it is possible to attribute the effects found for the two continua with [t] solely to the bias against the other consonant: against [p] in the *spVp* vs *spVt* condition (rather than against [t] in *stVp* vs *stVt* condition) and against [k] in *skVk* vs *skVt* condition (rather than against [t] in *stVk* vs *stVt* condition). Indeed, the size effects suggest that this interpretation is not unreasonable. The two continua involving [t] *together* only produce a difference in cross-over points that is less (1.04) than the difference achieved with the [p]-[k] continuum by itself (1.4). This might lead one to hypothesize that there is no bias against [t], and that [p] and [k] each have a bias of roughly half a scale point. Coetzee fends off this interpretation by means of a pairwise comparison of the subjects' difference scores in each of the three continua, so as to show that the difference scores for the [k]-[p] continuum are not significantly different from those obtained with the [k]-[t] continuum and the [p]-[t] continuum. It is entirely legitimate to conclude, as Popper would have us, that it has not been shown that [t] is treated differently from [k] or [p]. However, we must *not* think that it has been shown that [t] is treated in the same manner as [k] and [p]. Experimental results that show non--significance between potential conditions may be used to justify decisions to pool data, but provide a shaky basis for conclusions about the possibly different roles that the non-significantly different conditions might have to play in the world. The second threat to an abstract grammatical interpretation of the OCP-effects is that knowledge of transitional probabilities need not be confined to contiguous phonemes. Coetzee fends off this threat by pointing out that non-contiguous transitional probabilities are part of what lexical neighbourhood indices are based on, and that since these indices did not significantly affect the results, we can rest assured that non-contiguous transitional probabilities had no effect either. Again, it just hasn't been shown that they did. Also, the test of this possible effect is so diffuse that I would have

appreciated a more explicit explanation of the relation between transitional probabilities and neighbourhood density indices. Another aspect that might have warranted more discussion is why the effect of the OCP wasn't shown on the basis of the recognition of the final consonants of words like *pipe, kick* and *tot*. It might have been less problematic to argue that any effects would be unrelated to the frequent occurrence of these patterns. In spite of the only partially parried threats to Coetzee's interpretation, the contribution is recommended for the discussion of the experimental procedure.

In Galician, the velar nasal occurs word-finally before pause and vowels to the exclusion of [m,n,ɲ]. The problem with this neutralization is that in some words the velar nasal exceptionally appears intervocalically, as in [uŋo,uŋa] 'one', presumably as a result of a lexical rule that treats the suffixes of indefinites as if they were word-initial, and moreover that in that word-internal position, the pronunciation of the velar nasal differs from its word-final pre-vocalic pronunciation. Sonia Colina and Manuel Díaz-Campos argue that the word-internal intervocalic [ŋ] is a geminate, and account for this configuration by, quite plausibly, assuming that a word-internal onset requirement in effect causes the velar nasal to have both coda and onset status. They show that their interpretation is realistic by measuring the duration of the velar nasal in coda position, as in *c'un curandeiro* 'with a folk healer' as well as onset position, as in *c'un amigolo* 'with a friend', and comparing these with the duration of the word-internal velar nasal, as in *algunha* 'some'. Indeed, the word-internal velar nasal is significantly longer. A question one might have is why it wasn't also shown that other word-internal nasals are *not* longer than in coda and onset position (e.g. *pano* 'cloth' vs. *un tren* 'a train', *unha nora* 'a daughter-in-law').

An intriguing data set that must be well known among Romance scholars, but was new to me, is elegantly dealt with by Clàudia Pons in an OT analysis. The data, from Balearic varieties of Catalan, and the analyses are too extensive for me to be able to summarize here. A non-trivial result is the explanation for the a-symmetric depalatalization of [ʃ,ɲ,ʎ] in word-final position in Sardinian Catalan: word-finally, [ɲ,ʎ] go to [n,l], but [ʃ] remains. The answer, foreshadowed in Pep Serra's 1996 Barcelona thesis, is that depalatalization is general before coronal consonants, but that the plural [s] acquires a vowel before it due to *SIBSIB, which militates against adjacent sibilants. The situation in the singulars is thus a reflection of the distribution that has arisen in the plurals (*anys* [ans], *cavalls* [kaˈvals], *aqueixos* [aˈkeʃus], hence *any* [an] 'year', *cavall* [kaˈval] 'horse', *aqueix* [aˈkeʃ] 'this', through output to output correspondence.

Laura Colantoni shows how sound changes are not oblivious to the wider system of oppositions in the language. Assibilated rhotics are a widespread feature of Argentinian Spanish, which are currently changing (back?) into trilled rhotics. This happens in tandem with the change of palatal glides into palato-alveolar fricatives, evidently so as not to end up with a system that

contrasts assibilated rhotics (apical post-alveolar fricatives) with assibilated palato-alveolars (laminal post-alveolar fricatives). While this type of opposition is not unheard of, it is typologically rare. Colantoni bases her work on her own earlier field work for the Linguistic Atlas of Hispanic America and the Linguistic Atlas of Argentina, and it is great to see how that work has led to these interesting insights into the course of language change. This paper's quality is further enhanced by extensive acoustic data and a fine-grained regional breakdown of frequencies of variants.

Carmen Lúcia Barreto Matzenauer and Ana Ruth Moresco Miranda tackle the complex lexical vowel alternations and neutralizations in Brazilian Portuguese. They do this in a constraint-based analysis which takes full cognizance of the representational framework associated with work by Clements. Thus, unlike OT constraints, their constraints may stipulate phonological actions, such as 'spread' or 'have double association'. The analysis would appear to work well.

This volume of papers presented at the first PaPI holds a promise for the future. It is beautifully produced, as appropriate to the style the Mouton de Gruyter's Phonology & Phonetics Series in which it appeared.

References

Cutler, A., Wales, R., Coopers, N., & Janssen, J. (2007) Dutch listeners' use of suprasegmental cues to English stress. *ICPhS16*.

Gussenhoven, C. (1999) On the limits of focus projection in English. In *Focus: Linguistic, Cognitive, and Computational Perspectives* (P. Bosch & R. van der Sandt, Editors), pp. 43-55. Cambridge: Cambridge University Press.

Gussenhoven, C. (2004) *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Kubozono, H. (2002) Temporal Neutralization in Japanese. In *Laboratory Phonology 7* (C. Gussenhoven & N. Warner, Editors), pp. 171-201. Berlin & New York: Mouton de Gruyter.

Ladd, D.R. (1996) *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, D.R. & Morton, R. (1997) The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics* 25: 313-342.

Swerts, M., E. Krahmer & Avesani, C. (2002) Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics*, 30: 629-654.