Welcome to the Workshop on Processing of Prosody across Languages and Varieties (ProsLang), held in Wellington, New Zealand, on 29th and 30th November 2018.

The workshop brings together research on cross-linguistic and cross-varietal commonalities and differences in: the role of prosody in signalling information structure, particularly in the activation and resolution of contrast and contrastive alternatives; the integration of prosody and morphosyntactic cues in speech comprehension, e.g. as cues to information structure; the role of prosody in the management and interpretation of discourse; prosodic structure as an organisational frame in speech production or perception; and links between prosodic structure and multimodal speech cues such as gesture.

This booklet contains the abstracts for the ProsLang workshop. The abstracts are presented in the order in which they are listed on the programme on the next two pages. A pdf of this booklet is available on the ProsLang website if you want to see the content in a larger format.

The ProsLang workshop gratefully acknowledges Victoria University of Wellington, the Marsden Fund managed by the Royal Society of New Zealand Te Apārangi, the Australasian Speech Science and Technology Association, the Association for Laboratory Phonology and Oxford University Press for their generous support for this workshop.

The organisers would like to thank the following people for their generous help in reviewing submissions for the workshop: Bettina Braun, Jason Brown, Jennifer Cole, Nicole Gotzner, Jianjing Kuang, Chilin Shih, Candide Simard, Petra Wagner, Catherine Watson.

Wellington, November 2018

Sasha Calhoun
Paul Warren
Janet Fletcher
Olcay Turk
Mengzhu Yan
Please note: all oral sessions are in Hunter Lecture Theatre 119 (HU LT119); poster sessions and morning and afternoon tea breaks are in the Hunter Common Room (HU202). For lunch suggestions see separate map of cafés etc. on campus.

**Thursday, November 29**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:45 – 9:00</td>
<td>Opening session</td>
</tr>
</tbody>
</table>
| 9:00 – 9:50 | Invited Speaker: Janet Fletcher  
*Intonation and contrastive focus marking strategies in two indigenous languages of Australia and Vanuatu* |
| 9:50 – 10:15 | Jason Brown and Karsten Koch  
*Prosodically driven focus movement in Urama* |
| 10:15 – 10:45 | Morning tea                                                               |
| 10:45 – 11:10 | Candide Simard and Aicha Belkadi  
*Are you by any chance a new referent NP in this narrative?* |
| 11:10 – 11:35 | Abiola Oyelere, Candide Simard and Anne Lacheret-Dujour  
*Prominence in the identification of the focus elements in Naija (Nigerian Pidgin)* |
| 11:35 – 12:00 | Olcay Turk  
*Coordination of gesture with prosodic and information structure in Turkish* |
| 12:00 – 12:25 | Francois Nemo, Gilles Cloiseau and Fanny Krimou  
*Prosodic marking of questions under discussions: Empirical and experimental approaches* |
| 12:25 – 13:35 | Lunch                                                                      |
| 13:35 – 14:25 | Invited Speaker: Anne Cutler  
*Sources and domains of variability in prosodic processing* |
| 14:25 – 14:50 | Hilary Wynne, Linda Wheeldon and Aditi Lahiri  
*The preparation of complex prosodic frames by L2 English speakers* |
| 14:50 – 15:15 | Timo Roettger, Michael Franke and Jennifer Cole  
*Listeners rationally adapt to unreliable intonation* |
| 15:15 – 15:45 | Afternoon tea                                                             |
| 15:45 – 16:10 | Hong Zhang  
*Phrasal effect on Thai tone realization* |
| 16:10 – 17:00 | Invited Speaker: Jennifer Cole  
*Conventionalization in the prosodic encoding of information structure: An information-theoretic approach* |
| 17:00 | Closing                                                                    |
| 18:30 | Dinner – Foxglove, 33 Queens Wharf (see map at end of booklet; weather-permitting, a walking bus will leave from the Hunter courtyard at 17:45) |
Friday, November 30

8:45 – 9:35 Invited speaker: Nicole Gotzner
*The semantic and pragmatic function of focal pitch accents in German and English*

9:35 – 10:00 Yan Sun and Chilin Shih
*Effects of semantic similarity and focus on the resolution of ambiguous sentences*

10:00 – 11:30 Morning tea & Posters

| 1. | Daniele Scanzi, Jérémy Zehr, Francesca Foppolo and Florian Schwarz - Only vs. clefts: The incremental processing of presupposed vs. entailed content |
| 2. | Shinobu Mizuguchi and Koichi Tateishi - Lexical accent and focal prominence in Japanese |
| 3. | Yue Cheng - Tone patterns of Du Fu’s pentasyllabic lines |
| 4. | Katharina Zahner and Bettina Braun - *F0 is not enough* - when museum does not activate musical in *Australian English* |
| 5. | Grażyna Demenko - Expressive speech as a key for human-computer communication |
| 6. | Anna Dannenberg - How prosody affects perceiving syntactic boundaries in spontaneous speech |
| 7. | Jun Xia and Xiaofang Cui - The prosodic constraints on the coordinate construction of incomplete names in mandarin Chinese |
| 8. | Seppo Kittilä and Sonja Dahlgren - Prosodic marking of certainty contrasts: the case of Finnish modal-evidential particles |
| 9. | Hong Zou - A prosodic approach to the binding area of Chinese reflexive "Ziji" |
| 10. | Michael Howard - The intonational phonology of Sāmoan questions |

11:30 – 11:55 Simon Wehrle, Timo B. Roettger and Martine Grice
*Exploring the dynamics of backchannel interpretation: The meandering mouse paradigm*

11:55 – 12:20 Jesin James and Catherine Watson
*The role of prosody and semantics in the perception of secondary emotions*

12:20 – 12:45 Angelika Hönemann and Petra Wagner
*Local speech rate in attitudinal speech in German*

12:45 – 13:55 Lunch

13:55 – 14:45 Invited Speaker: Bettina Braun
*The role of pitch accent type and focus-sensitive particles in the activation of contrastive alternatives*

14:45 – 15:10 Mengzhu Yan and Sasha Calhoun
*What primes alternatives? Investigating syntactic and prosodic focus priming of alternatives in English and Mandarin Chinese*

15:10 – 15:40 Afternoon tea

15:40 – 16:05 Isabelle Franz, Gerrit Kentner, Luisa Bernius and Frank Domahs
*The influence of rhythm and animacy on word order in three different age groups*

16:05 – 16:30 Ivan Yuen, Nan Xu Rattanasone and Katherine Demuth
*Six-year-olds use pitch, not duration, to mark focus in Australian English*

16:30 – 17:10 Discussion

17:10 Closing
Intonation and contrastive focus marking strategies in two indigenous languages of Australia and Vanuatu

Janet Fletcher
University of Melbourne
janetf@unimelb.edu.au

In the last ten years or more, there has been an expansion of prosodic analyses of less-well studied languages (e.g. Steindel-Burdin 2015, Jun 2015, Himmelmann and Kaufman, in press). Compared to well-studied European and Asian languages, only a handful of investigations have examined the interaction between prosody and information structure in Australian and Oceanic languages with some notable exceptions (e.g. Calhoun 2015 for Samoan; Simard 2010 for Jaminjung, Fletcher & Butcher 2014 for an overview of Australian languages). It is generally accepted that tonal variation in languages like English is influenced by a combination of information structure and pragmatics. However, the phonological intonational devices that languages use to contrast informational versus neutral focus are known to vary. These might include combinations of the following: manipulations of phrase-level pitch range (incorporating pitch level and pitch span after Ladd 2008), intonational and prosodic phrasing, and intonational prominence, including the use of different types of pitch accents for contrastive emphasis. Languages can also de-accent material (reducing the number of pitch-accents in a phrase) and/or de-phrase non-focal material (reducing the number of intonational constituents) to promote a particular kind of discourse interpretation (after Jun 2015). There is a general observation that languages with diverse prosodic characteristics like Hindi, Mandarin, or Chichewa also tend to show post-focal pitch range compression, with prosodic phrasing, pitch accents, or lexical tonal contrasts left intact (e.g. in the case of lexical tone languages i.e. Mandarin and Chichewa).

Languages can use syntactic means to realise informational structure categories like topicalisation and contrastive focus. These devices include left dislocation or syntactic fronting of the constituent under focus. This has been extensively noted in both Australian and Oceanic languages (e.g. Vera’a, Schnell 2018, Nafsan, Thieberger 2006, Garrwa, Mushin 2005, Warlpiri, Simpson 2007, Mawng Singer 2016). It has also been suggested that intonation plays a lesser role in the realisation of semantic (as opposed to west-Germanic languages) with patterns of prosodic variation primarily the result of positional factors. In other words, if a language promotes left dislocation as a topicalisation or contrastive focus-marking strategy, the resulting prosodic patterns are because the item under focus is in initial position in a discourse segment. By contrast, others have suggested that so-called free-word order languages employ intonational devices (e.g. Zimmerman et al. 2011) implying that there is a deliberate prosodic strategy to place a constituent in focus. In a related vein, recent explorations of the complex interplay between prosody, pragmatics, and syntax in Samoan, suggest prosodically driven syntactic fronting is an important feature of information structure realisation in this language (Calhoun 2015).

In this paper the interaction between prosody and focus realisation strategies are compared in two languages: Mawng, spoken on Goulburn Island, Northern Territory, and Nafsan, a Southern Oceanic language spoken on the island of Efate in Vanuatu. Mawng has approximately 500 speakers, and Nafsan around 5000. Both languages have previously been analysed prosodically as stress languages although this is still under investigation (e.g. Billington, Fletcher, Thieberger under review). Nafsan has preferred SVO word order although object fronting is used in cases of topicalisation (after Thieberger 2006). Like many Australian languages, Mawng shows a wide range of different word orders (Singer 2016). Syntactic fronting and left dislocation are also observed. Our corpus consists of a series of controlled laboratory-phonology type speech experiments that were designed to explore prosodic realisation of neutral and contrastive focus on nouns that were subjects or objects in mini-dialogues. Four female talkers were recorded for Mawng, and eight talkers (five males and three females) for Nafsan. Word order was deliberately manipulated in the experiments to elicit contrastive focus on initial noun subjects and objects in the final utterance of the dialogue.

In contexts of contrastive focus, all speakers (Nafsan and Mawng) produce fronted focal elements with a major pitch movement associated with the focused noun (subject or object). The focused noun is also realised with a wider pitch span than the same token in non-focal contexts. They are always realised in their own prosodic phrase, and are often prosodically left-dislocated, particularly in Mawng, supporting previous findings (Fletcher, Stoakes, Loakes, Singer 2016). There is no evidence in either language that there is a phonologically different kind of focal pitch accent from a “regular” pitch accent. Post-focal material in Mawng is almost always produced in a relatively compressed pitch range with no evidence of post-focal de-accentuation. However, in the Nafsan data there is evidence of de-phrasing of non-focal nouns suggesting prosodic phrasing patterns similar to Korean, for example, which we do not observe in Mawng. Both languages exhibit edge-marking prosodic patterns that are amplified in contrastive focus contexts. The implication of these findings is considered in relation to prevailing models of prosody and information structure and current models of prosodic typology (e.g. Jun 2015, Himmelmann and Kaufman, in press).
Acknowledgements
The research presented in this paper was made possible by funding from the Australian Research Council through the Discovery Project scheme (DP110100938) and the Centre of Excellence for the Dynamics of Language (CoE140100041). I would like to acknowledge the contribution of my research collaborators to this research: Rosey Billington, Ruth Singer, Nick Thieberger, Debbie Loakes and Hywel Stoakes.

References
Singer, R. 2016. The dynamics of nominal classification. Pacific Linguistics
Prosodically-Driven Focus Movement in Urama
Jason Brown and Karsten Koch
University of Auckland and Alberta Education

We present data from Urama (Kiwaian) to illustrate how focus-driven movement provides evidence for phonological structure being accessible to syntactic operations prior to spell-out. We show that Focus-Prominence is satisfied only when a focused phrase can surface in the nuclear stress position -- in other words, unlike in Indo-European languages, default prosody trumps focus prosody. Finally, we show that focus-marking considerations can override some syntactic effects. We conclude that only phonological elements that are linked to discourse (as through focus-marking or D-linking) may be visible to the syntax prior to spell-out.

Many syntactic approaches to the syntax-phonology interface order syntactic operations before phonological ones [1,2], or have them outrank phonological constraints [3]. On the other hand, a phonological approach to focus argues that its only role is to assign prominence in prosody and that it has no syntactic consequences [4]. Under such an approach, focus-related movement could be purely phonological, but, if phonological movement is permitted at all, it should not have consequences for syntactic and semantic interpretation [5]. Data from Urama challenges these positions, where focus-related movement is conceptualized as an interface issue, dependent on the interleaving of both phonological and syntactic principles [6,7,8,9].

Urama [10] has a neutral SOV word order. Nuclear stress (shown by underlining), which is assumed to be indicated by peak F0, is preverbal, thus falling on the object in the canonical transitive order. In order to receive the main stress, focused elements appear in preverbal position. This is exemplified in (1) with the use of beha, a focus-associated particle. In (1), beha associates with the focused subject John in pre-verbal position, but is illicit in its (canonical) non-preverbal position (2). The appearance of a phrase in the preverbal focus position can thus force the scrambling of non-focused arguments out of preverbal position, as in (1). This ‘altruistic’ movement also holds for wh-phrases (taken to be inherently focused), as in (3). The standard approach to such movement is to assume that there is a high-ranking constraint demanding that focused elements are stressed (e.g. Focus Prominence in [9]). However, in order for such altruistic scrambling to occur, the syntax must have access to a representation that includes stress: (2) is ruled out not for syntactic reasons, but because the phonology is unable to assign stress to focused John in initial position. Scrambling of the object occurs in (1) so that stress can be assigned to the focused subject.

One solution is to propose that focus movement in Urama is purely phonological. However, this is also not the case. Focus-related scrambling is restricted by binding conditions. In response to the question ‘Who saw themselves?’, a possible answer with an anaphor would be the canonical (4), but not (5), where the non-focused anaphor scrambles leftward, parallel to focus-related scrambling in (1). Thus, focus movement is subject to syntactic constraints. Moreover, (4) shows that both syntax and default prosody can override Focus-Prominence, since focused Ginau does not receive nuclear stress at all here (unlike in English).

In contrast to anaphor binding, focus movement readily violates superiority in multiple wh-constructions. In these contexts, a focused wh-phrase can be focused and placed in preverbal position (6). This shows that different syntactic principles have different rankings with respect to focus-related movement operations. Superiority, unlike anaphor binding, can be violated to satisfy Focus-Prominence (similar to how D-linking in English can override superiority effects in which questions). We suggest that only discourse linked phonological constraints can override syntactic ones. We conclude by considering how phonological and syntactic principles in Urama are ranked with respect to each other in order to account for the focus system of the language.
(1) Go’ota=i John beha ro iho ka coconut=DEF John only NOM eat PRS
‘Only John is eating the coconuts.’

(2) *John beha ro go’ota-i iho ka

(3) Mevia uho=i hotu ro vema’ai ra?
Mevia fish=DEF who NOM gave INT
‘Who gave the fish to Mevia?’

(4) Ginau ro himiha ereve’a
Ginau NOM self see
‘Ginau saw himself.’

(5) *Himiha Ginau ro ereve’a.
self Ginau NOM see

(6) Na'u wotu ro eve’a
what who NOM see
‘WHO saw what?’ (compare with lit. tans. ‘What did WHO see?’)

Are you by any chance a new referent NP in this narrative?
Candide Simard, Aicha Belkadi
University of the South Pacific, School of Oriental and African Studies

This talk investigates the prosodic marking of referents in longer texts — personal anecdotes and mythological stories — in Jaminjung, a language of Australia, with a view to determining the convergence of phonetic prominence with prominence at other levels of linguistic structure, notably in syntax and semantics, where it usually refers to the higher rank of an element in a hierarchy. We make particular reference to the accessibility (or givenness) hierarchy: new > accessible > given (Chafe 1976, 1994; Gundel et al. 1993). It is assumed that as a discourse proceeds, information varies in its overall level of activation (see Wagner and Watson 2010 for an overview), and that more highly activated information tends to be highly accessible and requires less activation ‘costs’ from a speaker, which are reflected in its prosodic marking (Prince 1981, Givón 1983, Bauman & Grice 2006).

In Jaminjung, core argument NPs are marked by case and cross-referenced by bound pronouns on the predicate. NPs, however, are frequently omitted. Schultze-Berndt (2016) found that only 1.5% of Intonation Units (IUs) contain both overt A and O arguments (IU) in a dataset comprising 12,700 IUs. Word order reflects Information Structure (IS), with Topics, when present, occurring at the left-edge of an IU and making their own prosodic phrase, followed by a Comment where the focussed constituent is placed first and signalled by a [fall] pitch target on its initial syllables (Simard 2010).

We classify all overt referent NPs in the selected texts on a given-new continuum according to a non-binary taxonomy (Bauman & Grice 2006), coding for accessibility, activation and identifiability, and apply Givón’s (1983, 1995) quantitative approach to measure the continuity of referents i.e. whether overtly manifested or semantically implied as verb arguments; we also code for the syntactic role of the referents. Finally, we conduct an instrumental analysis of the prosodic correlates of the syllables of the referents, comparing them to neighbouring syllables, to quantify how they stand out (prominence): by bearing a specific pitch target, a greater pitch excursion, a slower speech rate (lengthening), or increased loudness. Our results show that the lesser degree of activation of a referent is consistently indicated by a stronger prominence, in accordance with cross-linguistic findings and predictions, and that this prominence is realised by a wider pitch excursion (pitch range). Our findings suggest a parallel encoding strategy for contextual givenness to that of the IS categories of topic or focus, and that this strategy entails gradience, which needs to be integrated in the current prosodic description schemas that comprise categorical distinctions only. We propose that, functionally, relative prosodic prominence in the NPs of Jaminjung narratives plays a crucial role in maintaining the coherence and cohesion of the narratives (Wichmann 2000).

This paper is a response to the programmatic call of Wagner et al. (2015) for a refreshed approach to the study of prosodic prominence, together with that of Himmelmann and Primus (2013) for a better understanding of the concept of prominence in linguistics. A comparison of how prominence is realised in the prosody of typologically diverse languages will reveal, in time, the language-specificity or universality of prominence and its correlates.
References

Prominence in the Identification of the Focus Elements in Naija (Nigerian Pidgin)

Abiola Oyelere¹, Anne Lacheret-Dujour¹ and Candide Simard²
biolaove2@gmail.com, anne.dujour.27@gmail.com, candide.simard@usp.ac.fj
1. Université Nanterre, Paris, France 2. University of the South Pacific, Fiji

This paper aims at examining the function of prosodic prominence in Naija, an extended pidgin spoken in Nigeria (West Africa) by the vast majority of its inhabitants. Prosodic prominence remains a complex notion that has received great attention among scholars this last decade, particularly in Romance and Germanic languages (for a recent review, see Wagner et al. 2015); in contrast, it has yet been little studied in African languages. Our work is based on the Rhapsodie model, developed for the intonosyntactic modelling of French, and its functional interpretation in terms of communication devices and shaping of the informational flow of speech (Lacheret-Dujour et al 2018).

Our study presents at least two major interests. First, it deals with a language that remains underexplored at the prosodic level. At the lexical level for example, Naija’s different varieties have been variously described as tonal, as stress, or as pitch accent languages (Mafemi 1971, Obilade 1976, Oyebade 1983, Faracas 1996, Elugbe & Omamor 1991). More than eighty-five percent (85%) of lexical items are of English origin (Faracas 1984). These English source words are used by Naija speakers who also speak any of about five hundred and fifty (550) different local tonal languages, which all have different tonal patterns (Blench 2012). The interaction of these two major sources results in a prosodic structure that is neither that of English nor of any of the local tonal languages. Despite and thanks to this complexity, Naija presents a fantastic opportunity to explore cross-language prosodic invariant features and cross-language interferences. Naija’s prosody beyond word level has yet never been described: the supra-lexical characteristics of prosody and its interplay with syntax in the marking of informational structure is unknown, and its place in prosodic typology remains unclear. We address the following questions: despite the diversity of its sources, is there a unified strategy to encode and perceive prominences in specific distributions, and how are these distributions correlated to informational structure?

The second major interest lies in the methodology of this study. We adapt a methodology, in terms of annotation, instrumentation, and automatic processing, developed for French, a typologically different language, to analyse the prosody of Naija, which makes it possible to highlight what is invariant and what is specific to prosodic measurement and modelling.

This study¹ is based on twelve five minutes monologues recorded from twelve speakers representing all the regions of Nigeria, presenting a balanced sampling between the Northern and Southern parts of the country. The data were orthographically and phonetically transcribed, then syllabified semi-automatically using the SPPAS tool². The resulting Praat textgrids were then annotated for prominence by three native speakers, all specially trained linguistics students at the University of Ibadan. Their points of convergence were noted by the researchers and automatically checked, making focused words within utterances visible. It then became possible to observe the most frequent syntactic units marked by prominences: their nature, their function and their distribution in the utterance.

Key words: Naija, information structure, prosodic typology, and focus

¹ Part of the NaijaSynCor project, sponsored by the French Agence Nationale de la Recherche (ANR).
² SPPAS (SPeech Phonetization Alignment and Syllabification) is a tool that produces automatic annotations which include utterance, word, syllabic and phonemic segmentations from a recorded speech sound and its transcription.
References
Coordination of Gesture with Prosodic and Information Structure in Turkish

Olcay Turk
Victoria University of Wellington

Gesture and speech are temporally coordinated. But what features of speech drive this coordination? Most studies investigated the temporal coordination between gesture and prosody [1]. However, gesture was suggested to originate from idea units being “novel departures of thought from a presupposed background” [2], i.e. the information structure (IS) of speech. IS and prosody are known to be linked [3], but a three-way alignment with gesture has never been systematically investigated. This study explores these alignments with gesture in Turkish.

Within the hierarchy of coordination proposed between prosodic and gestural structure, most studies focus on timing relationships between the smallest units, i.e. pitch accents and apexes [4],[5]. This study extends this investigation to larger units, i.e. intermediate phrase(ip)/intonational phrase(IP) and gesture phrase(GP). We also investigate IS-gesture relationship as part of a possible three-way alignment process. IS describes the salience and the organization of information in a discourse [2]. This study analyses the temporal alignment of gesture units with IS units that relate utterances to previous discourse (topic), and carry new information (focus). Based on these definitions, an alignment in time of these IS units with gestural units is expected because speech and gesture stem from "growth points" with informative goals [6]. Therefore, gesture should reflect qualities of the whole and integrate IS features due to marking “newsworthy” information within immediate discourse [7].

Our data comprises narratives from 4 monolingual Turkish speakers who watched 5 pairs of videos and recounted what they saw to a confederate listener. The design elicited different IS constructions. The recorded narratives were annotated using ToBI for prosody [8], McNeill’s guidelines for gesture [2] and the QUIS framework for IS [9].

Preliminary results show consistent gesture-prosody alignment, but with differences from previous studies because of the prosodic characteristics of Turkish. Apexes align with pitch accents; however, not all prosodic words have pitch accents in Turkish [10]. In such cases, they align with prosodic word-initial tones. Unlike English [1], single GPs do not align with single ips because of relatively short duration of ips in Turkish as an agglutinative language. However, onsets and offsets of these units are found to be sensitive to each other, that is, they start and end within 340ms of each other. Single IPs are also found to not align with single GPs and their onsets/offsets are not sensitive to each other. For gesture-IS alignment, contrary to the assumption that gesture would accompany foci as the most newsworthy information, topics are also found to be accompanied by gestures just as frequently. In terms of temporal coordination, topics/foci are nested within GPs, in that topics/foci align with a combination of core gesture phases that make up a GP. A further investigation revealed that the delay in the onset alignment is motivated by the time pressure caused by the prosodic phrasing of a previous utterance that is not gestured. The delay in the offset alignment seems to arise by gesture type. Metaphorical gestures tend to end earlier than iconics and deictics, which may imply that their function is more speaker-oriented (aiding lexical retrieval) rather than audience-oriented (conveying meaning).

These results support that prosodic and gestural structure systematically align, but with variation based on languages' intonational characteristics. This study is one of first to link IS and gesture. It shows that there is evidence for a gesture-IS coordination adapted according to prosodic constraints, making a three-way alignment process possible. This suggests three modalities interact with each other as they reach their surface forms, having implications for speech production models.
Prosodic Marking of Questions Under Discussions:
Empirical and Experimental Approaches
François Nemo, Gilles Cloiseau & Fanny Krimou (University of Orleans, UMR CNRS 7270)

The relation between prosody and questions under discussions (QUDs) has been mostly discussed in the study of focus and in relation with the prosodic building of the alternative sets which frame the interpretation of isolated utterance (e.g. Roberts, 1996)

Our aim in this context will be to show first that in more empirical and corpus-based studies of the relationship between prosodic marking and questions under discussion, it is necessary to:
- consider the relation between contributions in the sense of Grice (1975) and Horn (1993) and QUDs;
- consider as an empirical reality the existence of a discrepancy between the alternatives which would be considered as potentially relevant when considering the sentence alone and the alternatives which appears to be relevant and prosodically driven in the study of authentic data;
- acknowledge the metacontributional nature of prosodic comments, including focus;
- study prosodic variation in contexts in which QUDs can be experimentally manipulated by the linguist.

As for the first issue, after recalling the shortcomings of the confusion between the two notions of contribution and utterance, we shall show that QUDs at contribution level must be distinguished from QUDs at sentence-level, even when contributions are reduced to a single (uttered) sentence.

As for the second issue, we shall show that in order to describe authentic data and allow the description of parameters which appear to be prosodically relevant, three steps are necessary. The first step is to enlarge the notion of alternative sets into a wider notion of alternative frame, which allows among other things the description of the scalar value of the alternatives. The second step is to combine the description of the information shared within the alternative set/frame by both alternatives (presupposition) with the information provided by the alternative frame, in order to obtain an information frame. The third step is to describe the role of prosody in the construction of this information frame by the interpreter.

As for the third issue, we shall detail various illustrations of the way prosodic contours can be used to frame the on-going contribution.

As for the fourth issue, we shall present an experimental manipulation of contributional QUDs in the study of classical prosodic focus (e.g. John went to NORWAY) in order to illustrate the difference between alternative frames and alternative sets and in order to describe the relation of both with the actual prosodic realization of the focus.

All issues but the last experiment will be discussed in association with the study of authentic data, whose prosodic and semantic/pragmatic characterization will be provided. The data considered will include a large corpus of syntactically diverse injunctive utterances, notably declarative, interrogative and imperative ones.

References


Sources and Domains of Variability in Prosodic Processing
Anne Cutler1,2

1MARCS Institute, Western Sydney University; 2Max Planck Institute for Psycholinguistics

The title of the ProsLang workshop encourages contributors to compare the processing of prosody across different sources: languages with varying prosodic phonology, or varieties of the same language which differ prosodically in some way. To this shortlist of sources of variability in processing prosody one could add individual talkers (wherever individual differences are sought, they tend to be found, so there are bound to be individual differences in the manner and extent of prosodic processing too!). Individual differences in speech perception is as yet a little-studied topic (but see [1]). Not even so long ago, this was also true of cross-language comparison insofar as the implications of language-specific phonological structure for processing were concerned, but ProsLang and other workshops, plus several recent special issues, testify to quite an explosion of new interest in this domain. Varieties have been compared in this respect too, though such comparisons so far have not so much offered principled contrasts between well-described inter-varietal structures, as last-ditch explanation when no other source of a processing difference can be identified.

The domains of prosodic realisation addressed in these comparisons have mainly been prosodic salience and prosodic phrasing. Although these have long been believed to be where universals of prosody are in principle to be found [2], it has also been noted that in practice, the availability of an effective non-prosodic cue (e.g., markers for informational salience [3]) may make prosodic realisation of the same information redundant and hence less likely, or competition for the same acoustic dimension between prosodic and non-prosodic expressions (e.g., intonation and lexical tones [4]) may reduce the scope for realisation of both, with the prosodic dimension more likely to be found on the losing end.

As new evidence has accumulated in the recent expansion of comparative prosodic processing research, however, it seems to be interestingly the case that such warnings may have been premature. This is particularly true in the case of salience, where listeners appear to be able to attend to separate and subtle cues wherever they are informative, even those cues that can in principle co-vary [5,6]. More scope then becomes available for the resolution of potential competition, by allowing different aspects realised in the same dimension to convey different streams of information [7]. A similar account can be given for processing of juncture and lexical-level structure [8-10]. A further interesting extension of such variability concerns L2 usage when the L1 pattern is in full or in part different; here the reported listener behaviours range from perception in an L2 outdoing perception in the L1 [11] through simple L2 failure to use cues that L1 listeners use [12] to L2 listeners failing to exploit in the L2 the same cues they use in L1 [13]!

This is not the only area in which a range of results patterns is observed. Although less attention has been accorded to cross-variety comparison, here too there has been growth, including reports of perceptual correspondence despite varietal divergence in realisations [14]. Varietal-based explanations have also been offered for failures to replicate earlier perceptual results found in another variety of the same language [15,16]. Sometimes this happens even when replication success has occurred across the same varieties in other perceptual tasks [17,18], or when replication failures across languages already exist [19,20]. More principled guidelines for processing predictions based on proposed source discrepancy would undoubtedly be of use.


The Preparation of Complex Prosodic Frames by L2 English Speakers
Authors: Hilary Wynne1, Linda Wheeldon2, Aditi Lahiri1
1University of Oxford, 2UiA University of Agder

Multiword utterances regularly exhibit word boundaries that are not necessarily sacrosanct. For example, function words in English often encliticise to adjacent words to form a single prosodic unit (e.g. *pint of milk* > *[pa.ɪnt ə ˈmɪlk]*). This mismatch in prosodic and lexical structure points to a process in speech production where features of connected speech must be prepared: that is, where *drink a* becomes *drinka*: the phonological encoding process. English compounds are often identified by special semantic, morphological, and phonological criteria (cf. Lieber, 2006). Aside from word stress, there are other prosodic qualities that distinguish compounds from phrases: previous psycholinguistic research has indicated that, despite containing two or more lexical units, English compounds are treated as single prosodic units in connected speech (Janssen et al., 2008; Jacobs & Dell, 2014). However, much of this evidence comes from native (L1) English tasks, and comparatively little is known about how non-native (L2) speakers encode compounds and phrases for the purposes of production.

To examine this, we conducted four naming tasks containing 80 English stimuli: noun-noun compounds, adjective-noun phrases, disyllabic initial-stressed words, and monosyllabic words (see Table 1). Two groups of speakers participated: 50 native British English speakers and 50 native Bengali speakers highly-fluent in English (English track in education). Experiments 1 and 2 presented the stimuli in delayed task conditions (Fig. 1) to both speaker groups, while Experiments 3 and 4 used online task conditions (Fig. 2). All target items were presented in the plural e.g. *dishcloths*, and the auditory prompts (e.g. *What are good?*) were as to encourage reduction and attachment of the auxiliary *are* in responses: e.g. *dishcloths*’re good. If compounds are indeed treated as single prosodic unit in connected speech, then it follows that auxiliaries should reduce and attach to them to form a larger, single prosodic unit.

Data were submitted to linear mixed effect models in which naming latency was modelled as a function of condition and frequency. There was no effect of any frequency measure (surface, cumulative or constituent), and an overall effect of condition on naming latencies (p <.001*). In the delayed tasks, latencies reflected the total number of prosodic units in the target sentence: the adjective-noun condition elicited significantly longer latencies in both L1 (400 ms, t=10.63*) and L2 speakers (448 ms, t=4.2*) than the compound condition, which showed no difference in either speaker group (all t’s < 2.00) compared to the other conditions. In the online task, however, speech latencies only reflected the complexity of the first prosodic unit. The results showed significantly shorter latencies for the adjective-noun condition (190 ms, t=-2.40* in the L1 speakers; 240 ms, t=-10.51* in L2) compared to the compound and disyllabic conditions, while the monosyllabic word condition elicited slightly longer latencies in both speaker groups (208 ms, t=-2.41* for L1, and 267 ms, t= -6.52* for L2).

In all four experiments, we found evidence that both L1 and L2 speakers of English treated compounds as single prosodic units, to which the reduced auxiliary *are* attached. While results from the delayed task confirms that compounds and phrases are being treated differently, the strongest evidence for how they are treated comes from the online task results. Here, the mean naming latencies for the compound and monomorphic word conditions reflect larger prosodic units containing the encliticised auxiliary (e.g. *dishcloths*’re, donkeys’re, ducks’re), while latencies for the phrasal condition only reflect the first prosodic unit (e.g. *drab in dark cloths*). Our findings lend support to the claim that it is in fact the prosodic structure (not the lexical or morphosyntactic structure) of the utterance that is dictating the arrangement of prosodic frames during phonological encoding and that L2 speakers, if able to access the prosodic structure of the language, differentiate between compounds and phrases in English.
Table 1: Example stimuli

<table>
<thead>
<tr>
<th>Noun-Noun Compounds</th>
<th>Adjective-Noun Phrases</th>
<th>Disyllabic Words</th>
<th>Monosyllabic Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>dishcloths</td>
<td>dark cloths</td>
<td>donkeys</td>
<td>ducks</td>
</tr>
<tr>
<td>bookshops</td>
<td>big shops</td>
<td>barrels</td>
<td>brooms</td>
</tr>
<tr>
<td>nightshirts</td>
<td>new shirts</td>
<td>nickels</td>
<td>nuns</td>
</tr>
</tbody>
</table>

Figure 1: Naming latencies for Delayed Production Tasks (in ms)

Figure 2: Naming latencies for Online Production Tasks (in ms)


Listeners rationally adapt to unreliable intonation
Timo B. Roettger1,2, Michael Franke3, Jennifer Cole1
1Northwestern University, 2University of Cologne, 3University of Osnabrück

Introduction: Intonation plays an integral role in comprehending spoken utterances. However, it also exhibits a large degree of variability, raising the question as to how a listener reliably uses intonation to infer what a speaker intends to communicate. Despite this uncertainty, listeners rapidly integrate intonational information to predictively map a given intonational cue onto respective speaker intentions [e.g. 1,2,3] and adapt intonation-based inferences by what appears to be quick adjustment of beliefs about speaker production likelihoods [4,5]. Although less focused on in the literature, the absence of a pitch accent can also be predictive [6]. Indeed, rational interpreters should not make a categorical difference between presence and absence of a cue but show behaviour merely as a function of differential production likelihoods. Against this background, we present mouse-tracking data pertaining to the incremental interpretation of present or absent intonational cues prior to lexically disambiguating material, with a between-subject manipulation of stochastic reliability of these cues.

Method: In a preregistered study, sixty German listeners were exposed to two different intonational contexts in a forced choice mouse tracking experiment [7]. As an answer to the question ‘Has the wuggy collected the violin then?’ (English translation of original German), they were exposed to either (1) ‘The wuggy then HAS collected the violin.’ or (2) ‘The wuggy then has collected the PEAR.’ In (1), a pitch accent on <has> indicates that the object has already been mentioned in the question (VERUM). In (2), a pitch accent on <pear> indicates that the speaker refers to a contrastive object (CONTRAST). Absence of a pitch accent on the auxiliary in (2) can be utilized as an informative cue to the contrastive nature of the referent. As a control condition, listeners were exposed to intonationally uninformative utterances, in which listeners had to wait for lexical disambiguation (LEXICAL). Subjects were randomly assigned to two groups. In the reliable speaker group (RS), the mapping of pitch accent position and referent resolution was always reliable (VERUM → given; CONTRAST → contrastive); in the unreliable speaker group (US), the mapping was crossed in 1/3 of trials, making the form-function mapping stochastically unreliable. Acoustic stimuli were resynthesized such that listeners could not utilize any prosodic information preceding the critical pitch accent (or absence thereof). Listeners had to click on visually presented pictures to indicate the intended referent during the unfolding auditory stimulus. The computer mouse trajectories (x,y coordinates) of their responses were recorded with a 100 Hz sampling rate. We present results for matching trials (i.e. discourse status and intonation match) and the turn-towards-target measurement (TTT), operationalized as the last point in time at which the trajectory horizontally turns towards the target (see Figure 1).

Results and discussion: Raw data and scripts can be assessed from our osf.io repository (osf.io/dnbuk). The results are shown in Figure 2. There was a significant effect of intonational context, such that VERUM elicited significantly earlier TTTs than CONTRAST (p<.001), which elicited earlier TTTs than LEXICAL (p<.001, see Figure 1). There was also a significant interaction of GROUP and VERUM, indicating substantially later TTTs in US (p<.001). CONTRAST was not significantly affected by US. These results conceptually replicate earlier findings, demonstrating the rapid integration of intonational cues during intention recognition [e.g. 1-3]. The acoustically early cue associated with VERUM allows listeners to infer the intended given referent before the lexical material becomes available. Listeners also use the absence of this cue (no accent on <has>) to anticipate reference resolution. We propose a formal model of rational rapid cue integration by linking the TTT measure to the listeners posterior odds over interpretation alternatives, given an incrementally unfolding utterance. Based on a few defensible qualitative constraints on beliefs about speaker production likelihoods (e.g., a pitch accent on the auxiliary is more likely when referring to the given referent than referring to the competitor) we show that observed differences between VERUM and CONTRAST are compatible with rational incremental interpretation.

We will further present a series of follow up experiments that replicate and extend these findings, investigating the interplay between perceptual salience of the probabilistic cue (presence vs. absence of pitch accent) and its position within the utterance (prenuclear vs. nuclear).

The present line of research thus contributes to our understanding of how listeners deal with the ubiquitous uncertainty in processing intonation. We suggest that listeners infer speaker intentions based on both bottom-up acoustic cues as well as on dynamically adaptable probabilistic expectations about likely intonational contours in a given context.
Figure 1. Horizontal cursor position of time- and space-normalized averaged trajectories for lexical disambiguation, verum focus, and contrastive focus in the reliable-speaker group (circles) and the unreliable-speaker group (triangles).

Figure 2. Estimates and 95%-Credible intervals for turn towards-target-measures across focus conditions (color) and speaker-groups (shape). Small semi-transparent dots indicate mean values for each subject; grey lines group subjects. Dashed line indicates the average acoustic onset of the referent.

Unreliable trials mainly affected the predictive value of verum focus, leading to a substantial increase in the TTT measure.

Thai has traditionally been analyzed as having a five-tone system, namely the three static tones: High, Mid and Low and two dynamic ones: Rising and Falling (Abramson, 1962). Interestingly, contour shapes of Thai tones appear to vary depending on its position within a phrase (Morén & Zsiga, 2006; Potisuk, Gandour & Harper, 1997). It has been reported that full contours are only realized at phrase final positions or in syllables articulated in isolation. Alternations in contour shape are expected in other phrasal positions, while the five-way contrast is still preserved (Gandour, 1974). Although much focus has been placed on the study of fully realized tones, relatively little is known with regard to the tone realization at non-final positions. We therefore perform a comprehensive corpus analysis of Thai tone across three phrasal positions: initial, middle and final, to examine phrasal effect on tone realization.

We look at a corpus of read speech collected from 50 Bangkok Thai speakers. The corpus contains a total of 6,000 read sentences with about 120 thousand total number of words. Tones in the corpus were labeled according to the King Ross Thai pronunciation dictionary. 2-D kernel density estimation pooled across the entire sample of F0 contours and Functional Principle Component Analysis (FPCA) (Aston et al, 2010) are performed to visualize tonal contour variation and extract latent dimensions that quantify the observed variation. Figure 1 shows an example set of density plots at phrase final position, and Figure 2 plots the 4 latent dimensions identified through FPCA. The dimensions can be interpreted as reflecting properties of pitch height, contour slope, curvature and target location. Coefficients attached to the latent dimensions for each example are estimated by linearly projecting the original F0 contours onto the new coordinate.

Our analysis suggests that tones utilize different aspects of tone space under different phrasal contexts in Thai. These aspects can be quantified as, in addition to the traditional dimensions of pitch height and slope, details about pitch target distribution, including both the number and shape of the target, as well as the timing of target realization. Considering variations across tone categories at fixed phrasal position, it should be acknowledged that the proposed quantification dimensions largely agree with previous analyses on tone realization at phrase final position, while important information on target realization within the population is further included. It demonstrates that the underlying distinctions among tone categories in this context span variations in the number and shape of pitch target, as well as the timing of target realization. Such variations are efficiently captured by projecting the original contours to latent dimensions defined through FPCA. Similar density distributions across tones are found in phrase middle and initial positions. They both differ from phrase final tones in the timing and shape of target realization, in all tone categories. Although low and rising tone show similar density distributions in both initial and middle positions, they are not identical in the latent dimension corresponding to the timing of target realization. Therefore it can be predicted that the two tones are more confusable when unstressed, which may be argued as evidence for the existence of certain degree of neutralization, but not complete reduction in tone space.

Results from this study for the first time recovered the complex underlying structure of F0 contours, which is efficiently captured by latent FPC dimensions. It has also been suggested that these latent dimensions are effective in quantifying nuanced contour variations resulted from higher level prosodic factors.
Figure 1. Density plots of tonal contours realized at phrase final position.

(a) High tone  (b) Mid tone  (c) Low tone

(d) Rising tone  (e) Falling tone

Figure 2: Shape of the four latent dimensions identified from FPCA.

References:
Conventionalization in the prosodic encoding of information structure: 
An information-theoretic approach
Jennifer Cole, Northwestern University

Early accounts of phrasal prominence [Bolinger 1972; Chafe 1974; Chomsky & Halle 1968; Gussenhoven 1983; Ladd 1980] point to informational criteria (related to focus and discourse-givenness) and structural criteria (related to position in the prosodic phrase) as determining which word(s) within a prosodic phrase are assigned phrasal prominence. Yet empirical evidence from recent studies calls for reconsideration of analyses that directly and deterministically link information structure (IS) meaning with phrasal prominence and/or pitch accent. The argument against traditional accounts of the prosodic encoding of IS meaning rests on findings from experiments on English, French, German, Hindi, Italian, Russian, Spanish and Berber that collectively show: (1) the relationship between IS categories and phonological pitch accents is not deterministic—e.g., in English and German, accent may occur on given as well as new words, and many accent types (e.g., L*, H*, L+H*) are attested for each information structure condition. (2) There is systematic acoustic prosodic enhancement associated with gradations in informativity across IS categories—e.g., focused and/or new words are acoustically enhanced relative to given words—in languages with typologically diverse prosodic systems. (3) The gradient effect of informativity on acoustic prosodic enhancement can be greater for phrase-final (nuclear) prominence compared to non-final prominences, an asymmetry that is reflected in an overall structural bias that privileges the phrase-final position in perceptual processing of phrasal prominence.

These findings are hard to reconcile in traditional accounts that directly and deterministically link prosodic phonological structures with IS meaning. I argue for an alternative approach, adopting an information-theoretic framework where predictability and conventionalization interact to shape systematic variation in prosodic expression and its association with IS meaning. The predictability of a word and its referent influences linguistic expression at the lexical, syntactic, phonological and phonetic levels [e.g., Aylett & Turk 2004; Levy & Jaeger 2007], to varying degrees across languages. Traditional IS distinctions, re-cast in in terms of predictability, influence a speaker’s choice of prosodic expression at the phonological and phonetic levels, and variation in the prosodic expression of predictability is potentially offset by expression via lexical or syntactic choices. The conventionalized pairing of acoustic prosodic marking and IS meaning varies across IS categories and across languages. For instance, in English, a phonological pitch accent may come to function as a pragmatic morpheme due to the highly conventionalized pairing of acoustic prosodic enhancement (e.g., a sharply rising pitch excursion) with a salient information status distinction (e.g., corrective focus), while less conventionalized associations result in probabilistic and phonetically gradient patterns in speech production. In other words, the prosodic encoding of information structure is phonological only in the most conventionalized cases. This information-theoretic account offers insight into observed structural and pragmatic biases in the perceptual processing of prosody, and reconciles the apparent conflict between recent experimental findings in diverse languages and traditional views of the prosody-IS relationship.
In this talk, I will discuss the role focal pitch accents play at the semantics/pragmatics interface. I will present a series of experiments demonstrating that - in German and English - focal pitch accents activate alternatives in online processing and help listeners identifying relevant alternatives in context. Furthermore, our experiments indicate that focus plays a crucial role in inference processing, serving as a strong cue that a pragmatic inference should be derived. I will conclude by discussing how different formal pragmatic models can incorporate the role of focus in implicature computation.
Effects of Semantic Similarity and Focus on the Resolution of Ambiguous Sentences
Yan Sun    Chilin Shih
Department of Linguistics, University of Illinois at Urbana-Champaign

Rooth (1985, 1992) proposes that focus indicates a set of alternatives under discussion, and may influence the truth condition of an expression. For example, the following sentences are identical to each other with the exception of focus position:

(1) a. Besides John, [MIKE] also knows Bob.
   b. Besides John, Mike also knows [BOB].

In both cases, John is in the set of alternatives evoked by focus, yet the truth condition of (1a) is “John knows Bob, and Mike also knows Bob”, i.e., John is the alternative to Mike, whereas for (1b), the truth condition is “Mike knows John, and Mike also knows Bob”, i.e., John is the alternative to Bob. It is widely agreed that the set of alternatives does not comprise all entities which could be substituted for the focused item. There are further constraints. For example, Cohen (1999) suggests that alternatives are expressions which share a minimal presupposition with the focused item, and Umbach (2004) proposes that alternatives have to be both independent of and similar to the focused expression. Based on the previous literature, this paper reports an off-line reading comprehension experiment that investigates the impact of ‘semantic similarity’, a potentially quantifiable index, between alternatives on speakers’ interpretation of sentences like (1) under different focus conditions.

The experimental items are of the form ‘Besides the/his/her N1, the/his/her N2 also V the/his/her N3’. Semantic similarity values of noun pairs were calculated using the built-in algorithms in NLTK (Natural Language Toolkit) 3.0, and semantic similarity patterns were manipulated by exchanging the position of N2 and N3 in each item so that either N2 or N3 is more similar to N1. As for focus, we manipulated the capitalization of different nouns to indicate which noun is in focus position (N2, N3 or neither) and instructed the participants to stress the capitalized word when they read the sentences out loud. Before the main experiment, a plausibility norming test was ran to ensure that the two interpretations of each experimental item are equally plausible. Based on the results of the norming test, 24 item sets, with six sentences in each (see the example set below), were used in the experiment and six test lists containing 24 experimental sentences and 72 fillers were constructed using a balanced Latin Square design. In the experiment, each sentence was shown on a computer screen, together with three pictures corresponding to the three nouns in the sentence. Each picture could be dragged and dropped into different boxes representing either the subject or the object of the verb. 66 English native speakers participated and were instructed to read out loud each sentence and, based on their interpretation of the sentence, to drag and drop the pictures of the subject(s) into the boxes on the upper row and object(s) into the lower one (see Figure 1 for an example of the experimental interface).

Participants’ interpretation of N1 in each sentence was labeled as either subject reading or object reading. Mixed-effect logistic regressions were ran to model the probability of participants’ subject reading. As expected, focus has a significant effect (ps<0.001) on people’s interpretation: participants were more likely to interpret N1 as the subject (i.e., the alternative to N2) when the focus is on N2, and more likely to interpret N1 as the object (i.e., the alternative to N3) when the focus is on N3. More interestingly, the results also suggest a significant effect of semantic similarity (ps<0.001): when focus condition is controlled, participants are more likely to interpret N1 as the subject when N2 is more similar to N1 than N3 is (see Figure 2). Furthermore, the greater the similarity between N1 and N2, contrasted with that between N1 and N3, the more likely participants will interpret N1 as the subject (see Figure 3). Another finding is that participants rely more on focus to resolve the ambiguity when the focus cue and the semantic similarity cue are in conflict with each other.
Sample item set (in this example, father is more similar to mother than brother is)
a. Besides her mother, her father also believes her brother.
b. Besides her mother, her brother also believes her father.
c. Besides her mother, her FATHER also believes her brother.
d. Besides her mother, her BROTHER also believes her father.
e. Besides her mother, her father also believes her BROTHER.
f. Besides her mother, her brother also believes her FATHER.

Figure 1. The Interface of the Reading Comprehension Experiment

Figure 2. Predicted probability of subject reading as a function of focus and similarity pattern

Figure 3. Predicted probability of subject reading as a function of the similarity differences between (N1,N2) and (N1,N3) under different focus conditions

References
**Only vs. clefts: the incremental processing of presupposed vs. entailed content**

Daniele Scanzi*, Jérémy Zehr*, Francesca Foppolo*, Florian Schwarz*

*University of Pennsylvania, *University of Milano-Bicocca

There’s an ongoing debate about the backgrounded vs. entailed status of the content associated with presupposition triggers like *only* and *also*. It has been claimed that a sentence containing *also* entails its prejacent, and presupposes that at least one alternative is true (Karttunen and Peters, 1979); a sentence with *only* entails the denial of its alternatives, with the prejacent being merely presupposed (Horn, 1969) or implicated (Horn, 1979; Rooth, 1985; Krifka, 1993). Several studies in the past tested *only* and *also* in a variety of tasks and found contrasting results (Foppolo & Marelli, 2017; Kim, 2008; Romoli, Khan, Sudo & Snedeker, 2015; Schwarz, 2015).

We implemented a Visual-World paradigm in which participants’ eye-movements were monitored during the processing of sentences like (1). The task was to identify the target character on the basis of a comment made by a game master in a card game, in which players join a ‘Chosen House’ as represented by the suit of the card they were dealt. By comparing *only*, cleft constructions and controls in which the trigger was substituted by a neutral command like *Look*, we aimed at testing the incremental integration of presupposed and entailed contents, that are reversed between the two triggers, as shown in (1):

(1) a. *Only Sarah* will join the Chosen House of spades  
Presupposition: Sarah has spades  
Assertion: exactly one person has spades

b. *It is Sarah* who will join the Chosen House of spades  
Presupposition: exactly one person has spades  
Assertion: Sarah has spades

c. *Look!* Sarah will join the Chosen House of spades

The scenarios always showed two male and two female characters next to their card (Fig.1-2). We designed the critical trials so that participants could use the exclusivity information of the trigger, along with the information about the target character’s gender, to identify the target prior to the full disambiguation point constituted by the final noun (*spades*). Crucially, of the two characters whose gender was a match (female in this case), only the target character (top-left) had a unique suit in the critical scenario (Fig.1). In the control trials, both of the target-gender-matching characters had a unique suit, thus preventing early identification of the target (Fig.2). If presupposed content is integrated earlier during incremental processing, we predicted an earlier convergence to the target character for clefts than for *only*, for the former but not the latter conveys exclusivity as a presupposition.

We tested 46 English and 49 Italian participants on parallel versions of the task. Each participant was either assigned to the cleft or *only* condition. Each experiment consisted of 48 experimental items, including 6 test items (1a-b), 6 control items (1c) and 36 filler items controlling for potential task-specific confounds.

We found a convergence toward the target prior disambiguation in the Only condition, but only for English. (Fig. 3, top line-Table 1). This result suggests that the exclusivity inference is derived incrementally to locate the target character, at least in English. This process seems to be delayed in the case of clefts, in which the same content (that someone has a uniquely identifiable card) is presupposed. This delay, however, might depend on the material/design used, so further research is needed to control for possible intervening factors that might have affected results on clefts in English, as well as the absence of an early effect in Italian.
Figure 1. Critical Scenario for the test sentences in (1).

Figure 2. Control Scenario for the test sentences in (1).

Figure 3. The top graphs represent English, the bottom graphs represent Italian. Grey vertical lines represent the mean of suit onset, which is the actual disambiguation point in all sentences. The plots show the target advantage over the same-gender-competitor in the predictable (blue) vs. unpredictable scenarios (red) for the three triggers: clefts (left graphs), look (middle graphs), only (right graphs).

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Err</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.7872</td>
<td>0.3979</td>
<td>-1.978</td>
</tr>
<tr>
<td>Trigger (Look vs. Only)</td>
<td>-0.7350</td>
<td>0.4305</td>
<td>-1.707</td>
</tr>
<tr>
<td>Scenario (Predictable vs. Unpredictable)</td>
<td>-1.9084</td>
<td>0.4916</td>
<td>-3.882</td>
</tr>
<tr>
<td>Trigger*Scenario</td>
<td>1.5784</td>
<td>0.6005</td>
<td>2.628</td>
</tr>
</tbody>
</table>

Table 1. Output of fixed effects for English in which we contrasted Only and the control condition (Look) in the two scenarios (Predictable-Fig.1 vs. Unpredictable-Fig2) and their interaction.
Lexical Accent and Focal Prominence in Japanese
Shinobu Mizuguchi$^1$ and Koichi Tateishi$^2$
$^1$Kobe University, $^2$Kobe College

Lee et al. 2015 ([1]) studied the production and perception of contrastive focus cross-linguistically; they used corrective number sequences as in (1) and measured prosodic cues such as duration, intensity and pitch in focal positions in various languages (cf. Table 1).

(1) A: Is Mary’s number 215-418-5623? B: No, the number is 215-417-5623.

They found that focus digits, compared to unfocused digits, exhibit greater duration, intensity and pitch in Mandarin Chinese and American English, but not in Korean and Tokyo Japanese. Another finding is that strong acoustic differentiation leads to higher accuracy in perception.

We tested their hypothesis with the telephone number materials produced by a native male speaker of Tokyo Japanese. Our participants (20F, 2M, mean age 20.45, SD 0.87) did unexpectedly well (cf. Table 2), contra our prediction, based on [1]. We will scrutinize why the identification ratio of Japanese contrastive focus is higher than expected, and also why it is not as high as those of English and Chinese in this paper.

Japanese is a pitch language and its accent is characterized by a falling H*-L bi-tonal contour. Japanese lexical items are divided into two groups: Accented (A) and Unaccented (U). Due to the accent rules of Japanese, the materials in our experiment come in three accentual phrases (AP) of \([\text{AP}_1\text{UAA}]\)\([\text{AP}_2\text{UAA}]\)\([\text{AP}_3\text{[UA][UA]}]\), regardless the lexical accent patterns of the numbers. Table 2 shows that the identification ratio of AP3 is the worst.

Shinya 2009 ([2]) claims that lexical accent has a strong impact on the perception of intonational prominence in Japanese; an A word (e.g. ‘a’ni ‘big brother’) shows a higher F0 peak than an U word (e.f. ‘ane ‘big sister) (Accentual Boost) and an A word lowers the F0 of the following word (Downstep). An A word is therefore perceived with greater prominence than a U word, and Accentual Boost and Downstep make the pitch difference between a sequence of two words the biggest in AU sequence (e.g. ‘a’oi ume ‘blue plum’), followed by AA and UU. The one in UA is negative, i.e. A is generally higher.

We conducted a production experiment with the stimuli consisting of two words which cover the four possible accentual sequences. With \([+F(ocus)]\) conditions added, 8 Tokyo Japanese speakers recorded 48 stimuli (2 sets of \([A+F]A\), AA, AA+[F], A+[F]U, AU, AU+[F], U+[F]A, UA, UA+[F], U+[F]U, UU, and UU+[F] and distractors) at a sound attenuated room.

Figure 1 shows the results. For the analysis, we follow Ishihara 2016 ([3]) and take the normalized F0-means of the six measurement points: the 1st F0-minimum, F0-maximum, the 2nd F0-minimum of Word 1 (L1-1, H1, L1-2), along with those of Word 2 (L2-1, H2, L2-2).

Our findings are that (i) focus boost is relatively small and a focused word is not always higher in F0 than a non-focused word (cf. AA+[F], AU+[F] and UU+[F]), and (ii) the baseline lowers, even in the environments of UU and UA, where Downstep does not occur by definition.

We conducted a perception experiment with the stimuli recorded in the production experiment, with the prediction that the focus identification ratio varies, due to the varied acoustic environments. We recruited 23 participants (F12, M11, Mean Age 19.65, SD 4.0) and Table 3 shows the result. Our prediction is borne out; the focus identification ratio varies from 57.6% to 91.4%.

The results of our experiments lead us to claim that F0 boost alone cannot explain the prominence perception in Japanese. Unlike English, Japanese F0 peaks need to be processed ‘relatively’ due to the lowering baseline. Along with F0 boost, downstep/compression plays a significant role to make focus prominent; the downstep in A+[F]U is significantly different from the one in AU (W=41, p=0.038 on Mann-Whitney Test). We need to consider the effects of language-specific acoustic features like Japanese lexical accent and their interaction with intonational prominence to account for prominence cross-linguistically.
Table 1. Results of production and perception experiments adapted from [1]

<table>
<thead>
<tr>
<th></th>
<th>South Kyongsang Korean</th>
<th>Seoul Korean</th>
<th>Tokyo Japanese</th>
<th>Suzhou Wu</th>
<th>Standard French</th>
<th>Mandarin Chinese</th>
<th>American English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production</td>
<td>median z-score values of focused digits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intensity</td>
<td>0.64</td>
<td>0.13</td>
<td>0.10</td>
<td>0.48</td>
<td>1.73</td>
<td>1.19</td>
<td>0.95</td>
</tr>
<tr>
<td>pitch</td>
<td>-0.26</td>
<td>0.24</td>
<td>-0.24</td>
<td>0.53</td>
<td>0.97</td>
<td>0.36</td>
<td>1.28</td>
</tr>
<tr>
<td>Perception</td>
<td>55.6%</td>
<td>44.6%</td>
<td>-</td>
<td>-</td>
<td>94.9%</td>
<td>97.3%</td>
<td></td>
</tr>
<tr>
<td>identification</td>
<td>rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Identification ratio of Japanese listeners in words bearing contrastive focus (percentage values)

<table>
<thead>
<tr>
<th></th>
<th>1(U)</th>
<th>2(A)</th>
<th>3(A)</th>
<th>4(U)</th>
<th>5(A)</th>
<th>6(A)</th>
<th>7(U)</th>
<th>8(A)</th>
<th>9(U)</th>
<th>10(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>100</td>
<td>88</td>
<td>91</td>
<td>97</td>
<td>98</td>
<td>92</td>
<td>65</td>
<td>91</td>
<td>82</td>
<td>58</td>
</tr>
</tbody>
</table>

Figure 1. Mean normalized F0 of four accent types

Table 3 Normalized Difference between peaks(-1to1) and Focus identification ratio (percentage values)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak1 - Peak2</td>
<td>0.706</td>
<td>0.561</td>
<td>0.374</td>
<td>0.297</td>
<td>-0.208</td>
<td>0.05</td>
<td>0.1</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification</td>
<td>ratio</td>
<td>77.2</td>
<td>68.5</td>
<td>64.1</td>
<td>57.6</td>
<td>91.4</td>
<td>78.2</td>
<td>76.1</td>
<td>65.2</td>
<td>(%)</td>
<td></td>
</tr>
</tbody>
</table>

N.B. After we submitted our abstract, we have found out that Lee et al. 2018 [4] tested their hypothesis on Japanese. Though their result is different from ours in detail, the general tendency that focus position in the target sequence is crucial in perception is the same with our finding.

**Tonal Patterns of Du Fu’s Pentasyllabic Lines:** an discussion on the potential levels of metrical structure of Chinese regulated verse

Yue Cheng  Peking University

Chinese regulated verse with certain tonal patterns is a typical style of Chinese traditional poem. There are two types of tones: even and non-even (oblique) which constitute the legal tonal patterns. The basic rule, “1-3-5 principles” (Chen 1979, Chen 1980), means the tones of the first, the third and the fifth syllables of a line can be free while the second, the forth and the sixth syllables should be distributed as “even-oblique-even” or “oblique-even-oblique” in a seven-syllable line. However, the principle had a little difficulty to explain the existence of violations.

We suppose there is a hierarchy of markedness constraints in terms of tonal patterns. We have tested 662 pentasyllabic poems (4976 lines) of Du Fu, a proficient poet in tonal patterns, in the period of Prime Tang Dynasty.

We find 22 patterns in the range of materials. The usage frequencies among them show significant differences. (See results in the second page)

According to the usage frequency, we can divide these patterns into three levels:

<table>
<thead>
<tr>
<th>Level</th>
<th>Usage Frequency</th>
<th>Feature</th>
<th>Member</th>
</tr>
</thead>
</table>
| Top Level     | High(Over 300)  | i. different tone value in even syllable of a line.  
ii. different tone value in the 3rd and 5th syllable of a line. | EEOOE; OOOEE  
OOEEO; EOEE  
EOEOO; OEEOO  
EEOOO |
| Normal Level  | Medium          | i. different tone value in even syllable of a line.  
ii. different tone values of the last 3 syllables  
iii. tend to near a certain pattern | EEOEO*; EEEOE  
EEOOO*; OOOEO  
EOSEO; OEOEO |
| Illegal level | Low (Less than 20) | i. the same tone values in even syllable of a line, or  
the same tone value of the last 3 syllables  
ii. mostly near a certain pattern | OEOOO; OEOEO  
EOOOO; OOOEO  
OOOOO; OOEEO  
EOEOO; OEOOE*  
EOEEE |

Ascendant patterns are used most frequently. But illegal ones are quite limited. The only 3 exceptions with a mark “*” are the results of historical influence.

Furthermore, we observe some lines with unusual word orders to fulfill the requirement of tonal patterns. That may indicate a prosodic-influenced syntax grammar.

Key words: Tonal patterns, Chinese regulated verse, Pentasyllabic Lines,  Du Fu
The Usage Frequency of Patterns

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Frequency</th>
<th>Pattern</th>
<th>Frequency</th>
<th>Pattern</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.EEOOE</td>
<td>1052(12)</td>
<td>8.EEOEO</td>
<td>184</td>
<td>15.OEOEO</td>
<td>16</td>
</tr>
<tr>
<td>2.OOOEE</td>
<td>671(16)</td>
<td>9.EEEOE</td>
<td>161(0)</td>
<td>16.EOOO</td>
<td>12</td>
</tr>
<tr>
<td>3.OEEOE</td>
<td>657</td>
<td>10.EEOO</td>
<td>146</td>
<td>17.OEEOO</td>
<td>10</td>
</tr>
<tr>
<td>4.EEOEE</td>
<td>568(13)</td>
<td>11.OEOE</td>
<td>69</td>
<td>18.OOOO</td>
<td>9</td>
</tr>
<tr>
<td>5.OEEEO</td>
<td>523</td>
<td>12.OEOE</td>
<td>42</td>
<td>19.OEEE</td>
<td>5(0)</td>
</tr>
<tr>
<td>6.EOEEO</td>
<td>413</td>
<td>13.OEEO</td>
<td>29(0)</td>
<td>20.OEOO</td>
<td>4</td>
</tr>
<tr>
<td>7.EEOOE</td>
<td>345</td>
<td>14.OOOO</td>
<td>16</td>
<td>21.OEOEE</td>
<td>1(1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.OEEE</td>
</tr>
</tbody>
</table>

Note:

The number in brackets are the frequencies used in a first line of a poem, and not included in the former number. Usually patterns ended with an E tend to show up in a line of even number (while these ended with an O can only exist in lines of odd number, unless they are rhymes), but they can also appear in the first line.

Reference (mentioned)


In intonation languages, pitch accent type signals pragmatic functions [1]. Acoustically, pitch accent types differ in the alignment of f0 peaks in regard to stressed syllables, which makes the position of f0 peak an unreliable cue to lexical stress [2]. Previously, Zahner, et al. [3] have shown that Australian English (AusE) listeners, when instructed to click on museum, more strongly activate the stress competitor musical, when the target museum is produced with an early-peak accent, i.e., f0 peak realized on first, unstressed syllable, compared to a medial-peak accent, i.e., f0 peak realized on second, stressed syllable. Hence, suprasegmental pitch accents affect lexical activation in English. Acoustic analyses of the targets (e.g., museum) in [3] showed that the first syllable was louder and had more vocal effort in early-peak than in medial-peak accents. It is hence unclear which cue (f0, intensity) caused the temporary activation of cohort competitors (e.g., musical) in [3]. We here used resynthesized stimuli to isolate the effect of f0.

We used the 64 frequency-matched cohort pairs from [3]. The cohort members were segmentally identical up to at least the first consonant of the second syllable, but differed in stress position (e.g., WSW museum/SWW musical); their first syllable contained non-reduced vowels. Each cohort pair was combined with two unrelated distractors to be presented on screen. In 32 of the 64 cohort pairs, the auditory target was one of the cohort members (in 16 experimental trial the WSW-word, in 16 distractort trials the SWW-word); in 32 fillers it was one of the unrelated distractors. Hence, in an experimental trial, listeners heard “The next word is museum” and clicked on the target; musical served as the stress competitor. Targets in [3] were recorded with a medial-peak (L+H* L-%) and an early-peak accent (H+L* L-%) and spliced into one production of the carrier. Here, we PSOLA-resynthesized these targets such that the f0 contours were swapped between conditions. 40 AusE participants (29 female, O=25.6 years, SD=7.8) were tested at the MARCS Institute, Sydney, using SR Eyelink1000.

Fixations to the 4 words on screen were extracted in 4ms steps (Fig. 1). Empirical logits of competitor fixations were analysed in a pooled analysis for the data in [3], henceforth Experiment “Natural”, and the current experiment (Experiment “Resynth”), Fig. 2 left and right panel, respectively. Intonation condition and Experiment were modelled as fixed factors, participants and items as crossed random factors in an lmer testing for interactions between the two factors in 7 consecutive 100ms-windows (shifted target onset to offset). During the processing of the segmentally ambiguous part, there was a significant interaction between the two factors for window 3 (1035-1135ms, β=0.72, SE=0.37, df=1172, t=2.0, p=0.05), with an effect of intonation condition in Experiment “Natural” (β=-0.65, SE=0.26, df=594, t=-2.55, p=0.01) and no effect in Experiment “Resynth” (p>0.8); there was a tendency in the same direction for window 2 (935-1035ms). The effect of intonation condition was clearly preserved after the segmental uniqueness point in Experiment “Natural” (win4-6: 1135-1435ms, all p=0.04), but in Experiment “Resynth” the difference in competitor fixations is only numerical (win4: p=0.52, win5: p=0.12, win6: p=0.07). The f0 cue alone, as present in Experiment “Resynth”, is hence not strong enough to change the percept of an unstressed syllable as stressed in AusE. Only the combination of f0 and increased intensity on unstressed initial syllables, as they occur in naturally produced stimuli, affects lexical activation [3]. Our results will be discussed in terms of the ability to use suprasegmentals in word recognition in English listeners [4-6] and the role of f0 in this process [7] in specific, and their implications for spoken word comprehension in general.
Fig 1. Evolution of fixations to target (WSW, dark blue line), competitor (SWW, red line) and the two distractors (light blue lines) in experimental trials in the two intonation conditions (early-peak condition upper panel, medial-peak condition lower panel). Acoustical landmarks (grey dashed vertical lines) are shifted by 200ms, the time it takes to launch a saccade [8].

Fig 2. Competitor fixation across intonation condition (early-peak, red vs. medial-peak condition, orange). Results from Zahner, et al. [3] (Experiment “Natural”) are displayed on the left-hand side; results of the current experiment (Experiment “Resynth”) on the right-hand side.

Expressive speech as a key for human-computer communication

Grażyna Demenko
Institute of Linguistics
Adam Mickiewicz University, Poznań
lin@amu.edu.pl

Key words: pitch accent modelling, expressivity, phonetic convergence, human computer interaction

The purpose of this research is to show that, analogically to human-human communication, in human-computer interaction based on speech it is necessary to model expressivity variation, especially prosody as a fundamental component of successful dialogue processing [1,2]. As a reference, a speech corpus derived from the authentic Poznan police database with the recordings of the 997 emergency phone calls was used for preliminary expressivity level detection [3]. Out of 20 000 recordings collected in the database a few hundred were chosen for acoustic evaluation, the basis for that selection being a perceptual assessment [4,5,6,7]. The material was divided into four groups representing the most unambiguous situational contexts: extreme expressivity (19 speakers), expressivity (31 speakers), neutral (40 speakers) and depressive speech (18 speakers). An automatic annotation of accent has been used for extraction of specific pitch patterns in expressive speech. It was based on $F_0$ changes on an accented and postaccented syllable and the parameters related to the register and $F_0$ changes range [8]. Rising tones begin with letter L (for a dynamic tone) or L_ (for a static tone). Falling tones begin with letter H (for a dynamic tone) or H_ (for a static tone). The first letter represents the course of $F_0$ on an accented vowel, the next letter stands for the course of $F_0$ on the vowel directly following the accented vowel. If the basic tone on an accented and postaccented vowel is close to even, then such a tone is marked as even F. If the $F_0$ course within the accented syllable can be approximated by a quadratic function, then such a tone change is marked with L&H. If the tone on the accented vowel initiates $F_0$ rise on a few subsequent syllables then such a tone is marked as ascending L*. If the tone on the accented vowel initiates $F_0$ fall on a few subsequent syllables then such a tone is marked as descending H*. The most frequent types of accent structures have been distinguished for Polish: 1) pitch height change at the beginning of the accented syllable: dynamic rising pitch: LH, LL_, LH_; dynamic falling pitch: HL,HH_, HL_; 2) pitch height change after the accented syllable: static-rising pitch: L_H, L_L, L_H_, static-falling pitch: H_L,H_H; 3) level pitch: F; 4) falling-rising/rising-falling pitch: L&H/H&L; 5) ascending/descending pitch: L*/H*. Statistics have shown that accent in expressive speech is most frequently realized through $F_0$ level changes on an accented and postaccented syllable and shift of $F_{min}$ (eg.[9]), especially in highly expressive/stressful speech. The results of synthesis of pitch accents realized by level changes will be evaluated in further research with a speech corpus created for modeling of phonetic convergence consisting of 13 hours of dialogues in expressive and neutral scenarios between 16 pairs of Polish students and it will be implemented in spoken dialogue system within an ongoing project supported by the Polish National Science Centre, project no.: 2014/14/M/HS2/00631 “Automatic analysis of phonetic convergence in speech technology systems”. 
How Prosody Affects Perceiving Syntactic Boundaries in Spontaneous Speech

Anna Dannenberg
University of Helsinki

In our research we analyse the prosodic structure of spontaneous speech and compare it with syntactic structures of the same data. We use a continuous wavelet transform (CWT) based tool to reveal prosodic patterns of spoken language [1]. These patterns, in turn, can be compared with the grammatical analysis to examine the relation of prosodic and syntactic structures in spoken language.

In our study on co-occurrences of prosodic and syntactic boundaries and units in spontaneous English speech [2], we used data from the Buckeye corpus [3]. Our corpus included spontaneous speech samples from informal interviews of five female speakers of American English, consisting of 10076 words.

In the fully automatic and unsupervised prosodic segmentation process, the weighted sum of normalized f0, energy and segmental durations was used as an input signal for the CWT. Prosodic boundaries were determined by tracking minima across scales in the resulting scalograms, resulting in 1700 prosodic boundaries in the data.

The syntactic segmentation was performed by 20 informants, all of them having good to excellent skills in English and a good command of grammatical structures and concepts of English language. Each sample was segmented by four informants independent of each other. The informants were asked to tag every syntactic sentence and clause boundary according to example demonstrations. They had no access to the spoken data but only the transcription, so they had to perform the segmentation task without help from any acoustic cues. The segmentation resulted in 1965 syntactic boundaries, 61% of which were tagged by all four informants and 88% by at least three of them, and 182 boundaries that the informants had failed to notice were added afterwards.

According to our results, syntactic sentence and clause boundaries correspond relatively well to the prosodic boundaries calculated by the CWT. Of all the boundaries marked in the data, 906 were co-occurrences of prosodic and syntactic ones. This is 53% of all the prosodic boundaries and 42% of the syntactic boundaries in the data.

Our study also provides evidence that prosody affects the syntactic interpretation of spontaneous speech, even when the audio speech data is not available and the prosodic cues therefore have to be deduced solely from the transcription. Of the syntactic boundaries tagged by the informants, almost 45% were accompanied by prosodic ones, but the same only pertained to 13% of the syntactic boundaries that the informants originally had not tagged. In other words, almost nine out of ten syntactic clause or sentence boundaries that the informants had failed to identify were also prosodically unemphasized by the speaker. On the other hand, even though the informants were told to only tag syntactic clause and sentence boundaries, about 3% of the tags were separating e.g. fillers or non-clausal phrases instead of syntactic clauses and sentences, and many of these boundaries falsely tagged as syntactic ones were situated in a prosodic boundary. Thus the informants, despite having a good knowledge of English grammar, had in a number of occasions erroneously interpreted prosodic boundaries as syntactic ones.

Our results thus indicate a notable connection between prosody and syntax in perceiving and interpreting spontaneous speech. The syntactic boundaries co-occurring with a prosodic one seem to be easier to recognize than those not accompanied by a prosodic boundary, even if the prosodic cues are not overtly present in the transcribed data. In addition, even grammatically highly educated informants occasionally mistake prosodic boundaries for syntactic ones.
References:
The prosodic constraints on the coordinate construction of uncomplete names in mandarin Chinese

Jun Xia Xiaofang Cui
(Shenyang Normal University, China)

The coordinate constructions of complete names in mandarin Chinese are freely used without any particular constraints. But the situation is different when the coordinate constructions contain one or more uncomplete names (only surname or without surname). There are at least three rules as below.

First, when the coordinate construction is followed with an appellation which is shared by each of the items in the construction, the syllable number of each item should not be less than 2. For example (S=surname, N=name, SN=surname + name, A=appellation):

欧阳和上官老师 * 史和金老师 (S+S+A)
Ms OUYANG and SHANGGUAN Ms SHI and JIN
晓芳和桂香老师 * 晓芳和香老师 (N+N+A)
Ms Xiaofang and Guixiang Ms Xiaofang and Xiang
张丽和晓娜老师 * 张丽和娜老师 (SN+N+A)
Ms ZHANG Li and Xiaona Ms ZHANG Li and Na
晓娜和张丽老师 * 娜和张丽老师 (N+SN+A)
Ms Xiaona and ZHANG Li Ms Na and ZHANG Li
张丽和欧阳老师 * 张丽和王老师 (SN+S+A)
Ms ZHANG Li and OUYANG Ms ZHANG Li and WANG
欧阳和张丽老师 * 王和张丽老师 (S+SN+A)
Ms OUYANG AND ZHANG Li Ms WANG and ZHANG Li

Second, when one item of the coordinate construction is of uncomplete form with 2 syllables, the other item/items, which may be of complete form or uncomplete form, should be 2-syllable too. For example:

晓娜和张丽老师 ? 晓娜和张丽华老师 (N+SN+A)
Ms Xiaona and ZHANG Li Ms Xiaona and ZHANG Lihua
欧阳和张丽老师 ? 欧阳和张丽华老师 (S+SN+A)
Ms OUYANG AND ZHANG Li Ms WANG and ZHANG Lihua

Third, when there is not any conjunction in between the two items of the coordinate constructions (meanwhile there are usually some quantifiers following), the syllable number of each item should be the same (both one or two). For example (Q=quantifier):

吕王二人 * 欧阳上官二人 (S+S+Q)
The two: LV and WANG The two: OUYANG and SHANGGUAN
? 吕欧阳二人 ? 欧阳吕二人 (S+S+Q)
The two: LV and OUYANG The two: OUYANG and LV

The prosodic structure of phrase construction in mandarin Chinese is quite sensitive to the syllable numbers of the words. Since the phenomena are rare in other languages, as we know, it will help both Chinese and world linguistic research if they could be explained with general prosodic theory.

Key Words: prosodic constraint; coordinate construction; syllable number; mandarin Chinese
Prosodic marking of certainty contrasts: the case of Finnish modal-evidential particles

Sonja Dahlgren & Seppo Kittilä (University of Helsinki)

Finnish lacks evidentiality as a grammaticalized category, but Finnish obligatorily codes mood. In addition to grammatical mood, mood and also evidentiality can be expressed by a range of particles, such as varmaan (‘probably’), varmasti/varmana (‘certainly’), and ehkä/kai (‘maybe’). All these particles are grammatically completely optional; however, pragmatically their use is often of the utmost importance.

The modal/evidential particles of Finnish can be divided into two groups based on whether they stress certainty or uncertainty, in addition to which further divisions are possible based on the degree of (un)certainty they express. Particles like varmasti/varmana/varmuudella (‘certainly’) and todellakin (‘really’) are related to certainty, while kai/ehkä (‘perhaps/maybe’) and varmaan (‘probably’), while originally a colloquial variant of varmasti, has come through semantic change to express uncertainty or possibility. These contrasts are very clear in affirmation, and, e.g., varmasti always expresses certainty, while varmaan is always related to possibility. Changes in stress/prosody have no consequences for the semantic contents of these particles, stress only emphasizes their basic meaning. This radically changes in negation. The differences between the particles are maintained when the particles are unstressed, as in Kalle ei varmasti tule (‘Kalle will definitely not come’) and Kalle ei varmaan tule (‘Kalle will probably not come’). Placing the stress on the particles, however, changes the meaning of the particles related to (un)certainly, thus Kalle ei varmaan tule (stress on the particle) also means ‘Kalle will definitely not come’, while the meaning of the certainty particles is maintained regardless of stress. In negation, the meaning of the uncertainty particles is thus largely determined by stress, activating a semantic contrast between affirmation and negation, in negation separating the certain from the uncertain.

In our talk, we will present and discuss Finnish modal/evidential particles and show that similar changes in meaning apply to all of them (also to, e.g., ehkä ‘maybe’ and todennäköisesti ‘likely’); particles of certainty maintain their meaning, while the meaning of the uncertainty/possibility particles may be modified by stress in negation. Moreover, we will also discuss the rationale behind these differences. First of all, our study has relevance to the differences between affirmation and negation. The differences between the particles may be neutralized in negation by stress (all of them expressing certainty), which reflects the marked nature of negation; also many other differences are neutralized in negation, as for example the expression of aspect in Finnish by object marking, which is confined to affirmation. Second, our study is relevant to modality and evidentiality from a phonetic viewpoint. Stress is intuitively related to certainty in that it is easier to emphasize something we have certainty of, or that we find important. This is evident in negation, but also in affirmation, stressing a particle of uncertainty makes it more certain in meaning. Third, iconicity plays a role here. Stressed words are perceptually more salient, and through a connection to the non-linguistic world it can be said that stressing certainty is easier than stressing uncertainty. For example, relating to deixis, things that are in our vicinity (certainty) can be pointed at, whereas things not present (uncertainty) cannot. Similarly, we can emphasize something we know, while this is less natural for something we are not certain of.
A Prosodic Approach to the Binding Area of Chinese Reflexive "Ziji"

Zou Hong

(Jinan University Guangzhou China)

Abstract: Chinese reflexive "ziji" has three main characteristics: first, it can be bound within a long distance, sometimes across a clause; second, it tends to be bound by subjects instead of objects; third, it will be blocked when agreement features do not match. Present research about "ziji" mainly focus on the point of views from syntax, semantics, pragmatics and cognitive area. It is, actually, prosody gives the most important light on the distinctive features of "ziji", that is "ziji" can be bound in the phonological utterance as the maximum area but the last choice for its antecedent, the phonological word as the minimum area but the first choice for its antecedent, the phonological phrase and the intonational phrase as the another two possibilities. In different prosodic domains there are different rules for binding "ziji", so the usage of "ziji", at same time, gives a check to the Chinese prosodic hierarchy construction. And the relative prominence of different prosodic construction can help to interpret the ambiguity that "Ziji" brings to the sentence involving it for its diverse domains.

Keys: reflexive, "ziji", prosodic hierarchy, binding

Contact way: telephone number 18820096682  e-mail phylliszou@126.com

1. Prosodic Hierarchy of Putonghua

Foot << Phonological Word << Phonological Phrase << International Phrase << Phonological Utterance

2. Domain of the Foot

i. The words in lexicon but no more than three syllables

   e.g. xiaohair momode huidao zijide fangjian

   (little boy silently go back himself+possessive mark room)

   guanshang men. (The little boy went back his room in silence.)

   (shut door)1

   The foot is not the binding domain of "ziji", because there are only two possible foots involving "ziji", one "ziji" as a single foot, another "ziji" and a light tone as a foot. Antecedent of "ziji" can never be a light tone word.

3. Prosodic Word

3.1 Domain of the Prosodic Word

i. Based on the words forming rule, foots join together to form phonological words.

ii. No monosyllable word remain as an individual phonological word.

Xiaohair momode huidao zijidefangjian guangshangmen.

3.2 Phonological Words as the Binding Domain of "ziji"

Phonological word is the minimal binding domain of "ziji" and the rule of binding is as following:

\[ X_{ziji} g \omega \] (If some element precede "ziji" in a phonological word, that element is the antecedent of "ziji".)

   e.g. taziji tiaozhile yizhong qishui (He made a kind of soda water by himself.)

   (he self make one kind soda water)

4. Phonological Phrase

4.1 Domain of the phonological phrase

i. Phonological words with verb, noun and adjective join with left side phonological words without verb, noun and adjective as phonological phrases.

ii. Phonological words with focus word join with left side phonological words without verb, noun, adjective and focus word

1 All example sentences come from reading model essays of Putonghua Proficiency Test.
as phonological phrases.

Here focus word refers to "ziji" for it always has a relative high pitch in a sentence.

4.2 Phonological Phrase as the Binding Domain of "Ziji"
"haishiwomenziji yaoyou xinxin" (We should still need self-confidence.)
(still ourselves should have confidence )

In this sentence, "haishiwomenziji" is a phonological phrase. The bind rule in phonological phrase is as the following:

\[ [...\text{Noun/Pronoun}_{g}\text{...ziji}_{g}]_{0} \] (In a single phonological phrase only noun or pronoun precedes "ziji" can be the antecedent of "ziji". And If any noun or pronoun precedes "ziji" in a single phonological phrase, it must be the antecedent of "ziji".)

5. The Intonational Phrase

5.1 The domain of intonational phrase

i All branching elements except those attached in IP are considered as one intonational phrase.

ii A phonological string separates to the other phonological strings with a relative long pause, and it is always marked with punctuation in writing.

5.2 The Intonational Phrase as the binding domain of "ziji" (cf. I feel ignorance of myself.)

"wo gandao zijuide wuzhi"
(I feel self+possessive mark ignorance)

"Wo" is the antecedent of "ziji". It is a first person pronoun. The bind rule in intonational phrase is as the following:

\[ [...\text{Noun/Pronoun}_{g}\text{...ziji}_{g}]_{1} \] (In a single intonational phrase only noun or pronoun precedes "ziji" can be the antecedent of "ziji". And If any noun or pronoun precedes "ziji" in a single intonational phrase, it can be the antecedent of "ziji".)

6. The Phonological Utterance

6.1 The domain of the phonological Utterance

i A surface sentence ending with a full mark.

ii A string of sentences spoken by one speaker are around one topic.

6.2 The phonological utterance as the binding domain of "ziji"
The bind rule in the phonological utterance is as the following:

\[ \text{SpecI}_{g}(overt)\text{...ziji}_{g}\text{...} U \]
\[ \text{SpecI}_{g}(covert)\text{...ziji}_{g}\text{...} U \]
\[ \text{...ziji}_{g}\text{...} \text{SpecI}_{g}\text{...} \]

7. The Order Hierarchy of the Application of Binding Domain Rules

the Phonological Word Domain Rules \(\ll\) the Phonological Phrase Domain Rule \(\ll\) the Intonational Phrase Domain Rule \(\ll\) the Phonological Utterance Domain Rule i \(\ll\) the Phonological Utterance Domain Rule ii \(\ll\) the Phonological Utterance Domain Rule iii

| the Phonological Word Domain Rules | \([X_{g}\text{ziji}_{g}]_{0}\) |
| the Phonological Phrase Domain Rule | \([...\text{Noun/Pronoun}_{g}\text{...ziji}_{g}]_{0}\) |
| the Intonational Phrase Domain Rule | \([...\text{Noun/Pronoun}_{g}\text{...ziji}_{g}]_{1}\) |
| the Phonological Utterance Domain Rule i | \([\text{SpecI}_{g}(overt)\text{...ziji}_{g}\text{...}]_{1} \) |
| the Phonological Utterance Domain Rule ii | \([\text{SpecI}_{g}(covert)\text{...ziji}_{g}\text{...}]_{1} \) |
| the Phonological Utterance Domain Rule iii | \([...\text{ziji}_{g}\text{...}]_{1} \text{SpecI}_{g}\text{...} \) |

References:
Cheng, L. L 1987 "On the prosodic hierarchy and tone sandhi in Mandarin", Toronto Working Papers in Linguistics (Department of Linguistics, University of Toronto) 7:24-52
The intonational phonology of Sāmoan questions

Author: Michael Howard, Victoria University of Wellington

Using the Autosegmental-Metrical (AM) framework (e.g. Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Ladd, 2008), the present study looks at the intonation of polar and pronominal questions in Sāmoan. It is shown that the two question types do not exhibit differing intonation patterns and that the way in which their question status is marked challenges current assumptions of AM theory. The data for the study came from two sources - recordings of pronominal questions collected as part of Calhoun's (2015) study of focus marking, and polar question recordings created specifically for this study. Two native speakers were recorded reading a list of polar questions which varied in length and the position of focus within the question. Comparing the sets of data, it was found that the key characteristic of both polar and pronominal questions was an H+L- phrase accent aligned with the stressed mora of the final word. In addition, a sustained high pitch plateau was observed beginning at the H- preceding the final ip and continuing until the high in the final H+L- phrase accent. Focus was found to have no effect on intonation. The results present an interesting case where phrase tones not only align with positions of stress, replacing whatever pitch accents may otherwise have been there, but also where they are the sole means of indicating question status. This runs contra to approaches in AM theory which argue questions are signalled by particular configurations on an utterance nucleus (e.g. Ladd, 2008). Alternative, compositional, analyses (Pierrehumbert & Hirschberg, 1990) also require some modification to their underlying pragmatics to account for the findings.

Key references


Exploring the Dynamics of Backchannel Interpretation: The Meandering Mouse Paradigm
Simon Wehrle\textsuperscript{1}, Timo B. Roettger\textsuperscript{2} & Martine Grice\textsuperscript{1}
\textsuperscript{1}IfI Phonetik, University of Cologne; \textsuperscript{2}Department of Linguistics, Northwestern University

Backchannels (BCs) are a ubiquitous feature of spoken communication, crucial for signalling affect, interest and understanding. Although BCs are produced and perceived effortlessly in countless interactions every day, exactly how these listener signals are realised and interpreted has received little attention to date. It has been claimed that atypical BCs as produced by e.g. second language learners can have serious effects on communication and can lead to misunderstandings and negative character attributions [1,2,3]. Although this is in line with anecdotal evidence, there is an almost complete lack of experimental data in support of such claims.

One notable exception is the experiment reported in [4], in which subjects heard different versions of a single backchannel token in context, differing in its prosodic realisation (falling, level, rising). Listeners then judged tokens for politeness or dominance on a Likert scale. Although the results are intriguing, they are limited by the fact that BCs are essentially, and exclusively, a part of interactive spoken communication. This means that hearing and rating individual tokens post-hoc leads to a degree of artificiality that severely limits the external validity of any such findings.

We endeavour to avoid such limitations by dynamically tracking listeners’ social perception when they are exposed to excerpts of conversation involving BCs. To that end, we use mouse-tracking, adapting the meandering mouse paradigm developed for use in the Cognitive Reflection test ([5]). Subjects in our experiment heard speech from one side of a German dialogue that contained five BCs (total duration \( \sim 20 \) sec.). Subjects were instructed to rate the speaker as (un)friendly, (in)attentive or (dis)interested as quickly as possible once playback had finished by clicking on one of two labels arranged on the left and right side of the screen, respectively. The results in [5] lead us to predict that even with an instruction to only act at the end of playback, participants will produce at least some mouse movements throughout stimulus presentation. These movements would reflect reactions to the various BC tokens and accumulatively reveal a listener’s judgement as it unfolds over time.

For this pilot study, we tested 11 subjects (7 f, 4 m). Each subject completed 36 trials. Each stimulus started with a falling “ja” (yes). The remaining four BCs were resynthesized to yield four conditions: 1) Rise (all BCs rising), 2) Fall (all falling), 3) Flat (all level) and 4) None (no BCs after the first; only background noise).

The Fall condition received the most positive ratings overall, closely followed by Rise. The None condition received slightly negative ratings and Flat received the most negative ratings.

Figure 1 shows the mean mouse trajectories across all subjects and rating categories. Tracking the time course of mouse movements reveals that in the Flat condition, rated the most negative, subjects already steadily gravitate towards the negative final rating after the second BC with a level intonation (see arrow at \( \sim 8 \) sec.). Further, we can see that not all intonation contours are equally appropriate for all types of BC. For instance, whilst after the “ja” BC at \( \sim 8 \) sec. there is a positive trend for Rise (see arrow), but not for Fall, this picture is reversed for the following “genau” BC (see arrow at \( \sim 14 \) sec.). This is in line with production data we have gathered from Map Tasks which show that “ja” is predominantly produced with a rising intonation, “genau” predominantly with a falling intonation.

The meandering mouse paradigm allows us to test the perceptual effects of BCs in a fine-grained, dynamic, real-time perception task. Our pilot data suggest that this is a promising tool and show that, as conversation unfolds, listeners integrate the prosodic form of BCs in real-time in order to accumulatively change their social perception of a speaker.
Figure 1. Mouse trajectories averaged across all 11 subjects. Time (in seconds) on the x-axis, horizontal cursor position on the y-axis. Positive values represent the half of the screen containing positive judgements (i.e. interested, friendly, attentive), negative values represent the half of the screen containing negative judgements. The dashed lines indicate the start times of the five backchannel tokens contained in the stimuli, the dotted line represents the end of audio playback.

Speakers use prosody and semantic cues to reinforce shared meanings to communicate emotions in speech [1-3]. With the advent of speaking robots used in Human Robot Interaction, there has been an increased interest in understanding emotional speech for recognising and synthesising emotions. Acoustic-perceptual studies have shown that people can accurately identify emotions in speech only based on the prosody component, independent of semantic cues [4, 5]. Also, emotions can be perceived only from prosody variations even while listening to foreign language speech [6, 7]. But, these studies [1-7] have looked into primary emotions, which are innate emotions for fast and reactive response (e.g.: angry, happy). Here, we look into the effect that prosody and semantic cues have on the perception of 5 primary emotions (angry, happy, neutral, sad, excited) and 5 secondary emotions (anxious, apologetic, enthusiastic, pensive, worried). Secondary emotions arise from higher cognitive processes, based on evaluating preferences over outcomes and expectations. 4 professional New Zealand English speakers recorded 2400 sentences in the 10 emotions. The sentences were constructed to a) to study the effect of prosody variation alone (by using context neutral sentences) and also b) to study the combined effect of prosody and semantic cues (by using emotionally coloured sentences). As the speakers had difficulty in eliciting the subtle secondary emotions without any emotionally coloured words, some emotion inducing methods (eg: [8]) were also employed. A perception test with 120 participants, with 60 each evaluating the primary and secondary emotions was conducted. 50 of them were first language NZE speakers (L1) and 70 were bilingual English speakers (L2). The emotion perception accuracy for the primary emotions was 70%, while the secondary emotion perception accuracy was only 40%. Further analysis separating the context neutral and emotionally coloured sentences showed that the perception accuracy of the emotionally coloured sentences alone was 64%. This shows that the overall perception accuracy was brought down by the difficulty in judging the context neutral sentences. To understand if the speakers produced the utterances differently in the context neutral and emotionally coloured cases, prosody analysis was conducted. Mean Fundamental frequency (F0) and Mean Intensity were identified as the most significant features for distinguishing the emotions. The features were analysed for differences among the context neutral and emotionally coloured sentences (F0 boxplot shown Figure 1). F0 and Intensity showed no consistent difference among the 2 sentence types. This means that the speakers could portray the emotions consistently with and without the emotionally coloured words. The difference in the perception accuracy between L1 and L2 speakers was also analysed (Table 1). For the primary emotions, there was not much difference between L1 and L2, which is in agreement with previous research [4, 5]. For the secondary emotions, the L2 speakers found it harder to accurately recognise the emotions from the context neutral sentences compared to L1 speakers. But the inclusion of emotionally coloured words increased the emotion perception accuracy of both the L1 and L2 to a similar level. In conclusion, listeners found it harder to recognise secondary emotions compared to primary emotions. The prosodic features have variation across emotions types, but they alone are not sufficient to identify secondary emotions. When the semantic cues are in alignment with the prosodic variations, the nuanced secondary emotions can be accurately identified. Also, we have noticed differences among L1 and L2 speakers in their perception of secondary emotions without semantic cues, while both groups have similar accuracies in the presence of both prosodic and semantic cues. As the secondary emotions are used by people when interacting socially, the knowledge that both prosody and semantic cues are needed to perceive these emotions with be crucial in developing social robots. The dialog modelling in robots should be in congruence with the prosody component to develop socially interacting robots.
Figure 1. Variation of Mean Fundamental frequency (F0) for Secondary emotions for sentences with and without semantic influence.

Table 1. Perception Accuracy Summary table

<table>
<thead>
<tr>
<th>Emotion Type</th>
<th>Sentence Type</th>
<th>L1/L2</th>
<th>Perception Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>All</td>
<td>All</td>
<td>69%</td>
</tr>
<tr>
<td>Primary</td>
<td>All</td>
<td>L1</td>
<td>70%</td>
</tr>
<tr>
<td>Primary</td>
<td>All</td>
<td>L2</td>
<td>68%</td>
</tr>
<tr>
<td>Secondary</td>
<td>Emotionally coloured</td>
<td>L1</td>
<td>54%</td>
</tr>
<tr>
<td>Secondary</td>
<td>Emotionally coloured</td>
<td>L2</td>
<td>67%</td>
</tr>
<tr>
<td>Secondary</td>
<td>Context Neutral</td>
<td>L1</td>
<td>41%</td>
</tr>
<tr>
<td>Secondary</td>
<td>Context Neutral</td>
<td>L2</td>
<td>33%</td>
</tr>
</tbody>
</table>

References:
Local speech rate in attitudinal speech in German
Angelika Höinemann\textsuperscript{1,2}, Petra Wagner\textsuperscript{1}
\textsuperscript{1}Bielefeld University, Faculty of Linguistics and Literary Studies, Germany
\textsuperscript{2}Beuth University of Applied Science, Berlin

Alongside with fundamental frequency and intensity modulation, speech rate is an elementary prosodic cue which reflects the current affective state of a person, e.g. a speaker’s emotion [1]. We assume that attitudinal behavior has a great impact on the acoustic characteristics and that attitudes can be grouped together based on similar speech features, e.g. the speech rate, to identify attitudinal expression in a social dialog. Own studies show that the probability to misinterpret attitudinal behavior increases with the similarity of acoustic characteristics between attitudes [3]. In previous work, sixteen attitudes (cf. Table 1) were performed by sixteen speakers (8m, 8f, median age: 31.5) and were recorded embedded in short dialogs based on a paradigm which was used before [1, 3, 6]. Our data corpus includes a total of 512 acoustic recordings (13 stimuli were initially removed but estimated with an imputation algorithm with an NRMSE of 6% [7]). The speech rate metric in this work is based on a linear combination of phone and syllable rate, or perceived local speech rate (PLSR). Speech perception studies show that phone and syllable rate perception is not tightly correlated, but correlate tightly with the PLSR approach \((r=0.91)\) [4, 5]. In this work, we computed the percentage PLSR and the percentage variation of the speech rate (\(\Delta\text{PLSR}\), difference between two PLSRs). We found that with an increase in the PLSR, the speed variation increases as well \((r=0.67)\).

Performing an expressive attitude has a high impact on the speech rate and the \(\Delta\text{PLSR}\) respectively, thus the more expressive an attitude, the lower the speech rate/variation and vice versa. In previous work we determined the level of emotional activation as a measure for the expressiveness of an attitude [2, 3], thus correlations were found between both PLSR \((r=0.87)\) and \(\Delta\text{PLSR} (r=0.69)\) and the activation level. For instance, the interrogative attitudes SURP and DOUB show much lower PLSR/\(\Delta\text{PLSR}\) but high activations level than the neutral interrogative attitude QUES, the same is found for the neutral declarative attitude DECL which yields the highest PLSR/\(\Delta\text{PLSR}\) in contrast to the other expressive attitudes, e.g. SEDU, IRRI, ADMI and IRON. Attitudes such as POLI, AUTH and ARRO yield similar speech rates as the neutral DECL attitude, however, these can be seen as emotionally controlled, thus having less activation.

As mentioned before, we assume that groups of attitudes are perceived by humans based on specific acoustic characteristics. A 2-cluster analysis of the PLSR/\(\Delta\text{PLSR}\) based on the Euclidean distance reflects this assumption. Attitudes with a high speech rate are pooled together in one cluster \#2, e.g. the neutral attitudes DECL and QUES as well as attitudes such as POLI, AUTH, ARRO and SINC. Thus, this second cluster contains less expressive attitudes with a high speech rate. Contrary, cluster \#1 contains the expressive attitudes with lower speech rate, e.g. DOUB and SURP as well as IRO. SEDU, IRRI, UNCE and CONT. The similarity matrix displays the cluster results of PLSR and \(\Delta\text{PLSR}\). The blue fields point out similar speech rates between the corresponding attitudes, the red ones indicate low similarity between the speech rates (cf. Figure 1). For the PLSR, cluster \#1 is approximatively displayed on the bottom left side and the cluster \#2 on the upper right side. Noticeable are the low similarities of the attitudes CONT and UNCE to other attitudes. The cluster of the \(\Delta\text{PLSR}\) differs slightly from the PLSR clustering. Still, the attitude CONT shows a similar variation in speech rate compared to the more expressive attitudes. Thus, it is associated with cluster \#1. OBVI, which shows more similarities to less expressive attitudes, is associated to cluster \#2 - unlike in PLSR. As illustrated by the similarity matrix for \(\Delta\text{PLSR}\), the variations of speech rate are not separated as clearly as in PLSR, but follow the same tendency.
### Table 1: Description of the sixteen attitudes, the corresponding short terms using in this work and the attitudinal PLSR and attitudinal ΔPLSR

<table>
<thead>
<tr>
<th>Attitudes</th>
<th>Short Terms</th>
<th>PLSR %</th>
<th>ΔPLSR %</th>
<th>Attitudes</th>
<th>Short Terms</th>
<th>PLSR %</th>
<th>ΔPLSR %</th>
</tr>
</thead>
<tbody>
<tr>
<td>admiration</td>
<td>ADMI</td>
<td>70,19</td>
<td>0,64</td>
<td>Obviousness</td>
<td>OBVI</td>
<td>74,14</td>
<td>0,70</td>
</tr>
<tr>
<td>arrogance</td>
<td>ARRO</td>
<td>81,10</td>
<td>0,71</td>
<td>Politeness</td>
<td>POLI</td>
<td>82,31</td>
<td>0,80</td>
</tr>
<tr>
<td>authority</td>
<td>AUTH</td>
<td>84,57</td>
<td>0,74</td>
<td>neutral declaritive question</td>
<td>QUES</td>
<td>82,67</td>
<td>0,76</td>
</tr>
<tr>
<td>contempt</td>
<td>CONT</td>
<td>78,76</td>
<td>0,73</td>
<td>Seduction</td>
<td>SEDU</td>
<td>69,34</td>
<td>0,59</td>
</tr>
<tr>
<td>neutral declaritive statement</td>
<td>DECL</td>
<td>87,05</td>
<td>0,83</td>
<td>Sincerity</td>
<td>SINC</td>
<td>78,77</td>
<td>0,71</td>
</tr>
<tr>
<td>doubt</td>
<td>DOUB</td>
<td>66,56</td>
<td>0,62</td>
<td>Surprise</td>
<td>SURP</td>
<td>66,83</td>
<td>0,65</td>
</tr>
<tr>
<td>irony</td>
<td>IRON</td>
<td>70,54</td>
<td>0,62</td>
<td>Uncertainty</td>
<td>UNCE</td>
<td>71,54</td>
<td>0,52</td>
</tr>
<tr>
<td>irritation</td>
<td>IRRI</td>
<td>66,75</td>
<td>0,61</td>
<td>walking-on-eggs</td>
<td>WOEG</td>
<td>72,42</td>
<td>0,64</td>
</tr>
</tbody>
</table>

**Figure 1:** Similarity matrix for attitudinal PLSR (left) and ΔPLSR (right) based on the Euclidean distance (blue: high similarity, red: low similarity)

[5] Pfitzinger, Hartmut R., Local speech rate perception in German speech, In ICPhS-14, 893-896, 1999
[7] Stekhoven, D.J., Bühlmann, P., MissForest—non-parametric missing value imputation for mixed-type data, Bioinformatics, Volume 28, Issue 1, 1 January 2012, Pages 112–118
The role of pitch accent type and focus-sensitive particles on the activation of contrastive alternatives
Bettina Braun
Department of Linguistics, University of Konstanz, Germany
Keywords: prosody focus, topic contrast, additive particles, eye-tracking, online processing

In the information-structure literature, focus is generally defined as a constituent that corresponds to the open proposition in the context. In the semantic literature, narrow focus is defined in terms of the presence of alternatives that are relevant for interpretation (e.g., Krifka, 2008; Rooth, 1992). This semantic definition is close to what many researchers associate with the notion of 'contrast'. In fact, the relation between the categories focus and contrast is not entirely clear (Krahmer & Swerts, 2001; Molnár, 2001; Repp, 2010). For some authors, the notion of contrast is synonymous to (narrow) focus (e.g., Bolinger, 1961; Jackendoff, 1972; Krifka, 2008; Lambrecht, 1994; Rooth, 1992; Wong & Diehl, 2003), for others contrast represents a specific kind of narrow focus, resulting in a distinction between contrastive focus and non-contrastive (presentational, newness or information) focus (e.g., Bartels & Kingston, 1994; Baumann, Grice, & Steindamm, 2006; Kiss, 1998; Pierrehumbert & Hirschberg, 1990; Selkirk, 2002; Sudhoff, 2010; Watson, Tanenhaus, & Gunlogson, 2008 for studies on German and English). Experimental evidence from online processing speaks in favour of the distinction between contrastive and non-contrastive focus, based on differences in pitch accent type (Braun, 2006; Braun & Tagliapietra, 2010; Chen, Den Os, & De Ruiter, 2007; Husband & Ferreira, 2012; Kügler & Gollrad, 2015; Watson et al., 2008). However, given the nature of the materials and the tasks, differences in processing cannot uniquely be traced back to prosody.

The first goal of this paper is to test whether contrastive interpretation is also available without prosodic information that is available before or after the (contrastively or non-contrastively) accentuated referential expression. We will manipulate prosody on utterance-initial constituents (narrow focus with a contrastive pitch accent -- nuclear L+H* --, with a non-contrastive pitch accent -- nuclear H+L* --, compared to a control condition with a prenuclear accent in a broad focus utterance -- prenuclear L+H*) and will use the visual-world paradigm, which allows to time-lock fixations to contrastive alternatives before later prosodic information becomes available. The second goal is to compare the processing of intonational contrast to the processing of lexical items that presuppose the presence of alternatives (Büring & Hartmann, 2001; König, 1991; Reis & Rosengren, 1997), preceding the utterances from before by the additive particle auch 'also'. The third goal is to investigate the processing of words that are prosodically marked as contrastive topics (Büring, 1997).

Taken together, nuclear L+H* (contrastive focus) and prenuclear L*+H (contrastive topic) trigger more looks to a word that is contrastively related to the subject (compared to the control condition). A narrow focus with a nuclear H+L* accent (non-contrastive focus) or utterances preceded by the additive particle auch do not trigger more looks to a contrastive alternative. Implications for theories of information structure and the processing of additive particles are discussed.

References


What primes alternatives? Investigating syntactic and prosodic focus priming of alternatives in English and Mandarin Chinese

Mengzhu Yan and Sasha Calhoun, Victoria University of Wellington

A growing body of psycholinguistic research shows that contrastive associates are available to listeners upon hearing a sentence with contrastive accent, even when the alternatives are not explicitly mentioned in the discourse. For example, *The WIFE bought the clock* not only tells us that the wife bought the clock, but also implies alternatives to the focus such that the husband could have bought the clock. Increasing psycholinguistic evidence shows that these unmentioned alternatives (e.g. *husband*) are activated for listeners in sentences with contrastive accents (e.g. *The WIFE bought the clock*) (e.g.[2, 3, 4, 5, 6]). It is largely assumed that the mechanism behind this is (contrastive) focus marking: following from Rooth’s [1] alternative semantics theory, contrastive accents mark focus, which indicates contextually-relevant alternatives. However, to our knowledge, no previous studies have investigated whether other focus marking mechanisms also prime alternatives, e.g. syntactic focus marking, in the absence of prosodic focus marking. Across languages, morphosyntactic means of marking focus are as common as prosodic, and if focus is the underlying mechanism this should be the case; but this has yet to be investigated. Further, to our knowledge, all of the studies in this area have been carried out on Germanic languages, which have very similar prosodic systems. In this paper, we report on two parallel cross-modal lexical priming studies, carried out in English and Mandarin Chinese, which address these issues. The studies look at whether alternatives are primed by syntactic as well as prosodic cues to focus in these languages.

We tested the role of prosodic and syntactic focus marking in priming alternatives and general associates in English and Mandarin Chinese in isolated utterances in parallel cross-modal priming experiments in each language. Focus conditions and probe types used in experiments are shown in Table 1. The Chinese results showed that participants responded faster to alternatives compared to general associates and controls when the word was marked with nuclear prominence and syntax (SclleftS condition). No significant differences in reaction times were found across all probe types in the syntactic focus marking only condition (ScleftO) and in the no focus condition (canonO), which shows that syntactic focus does not prime alternatives. This might suggest that different types of focus marking are not equally efficient for the listener to infer alternatives. In addition, no facilitation for noncontrastive associates was found across all focus conditions.

With regard to the English results, this study is still ongoing; however, we predict a different pattern of results as syntactic focus marking carries more weight in English than in Chinese. We expected priming effects from both syntactic and prosodic focus as well as an additive effect when the both types of marking are on the same word. We anticipated a close relationship between weighting of focus-markings and ability to prime alternatives when comparing the two languages.

This research extends previous findings that prosodic focus marking primes alternatives out of context over a reasonably long time course; replicating this for English, and showing this for the first time in Mandarin Chinese. The research is also the first investigation of whether and how syntactic focus marking primes alternatives given its importance in signalling focus. Our findings suggest different focus marking mechanisms do not prime alternatives equally, but rather there is a crucial link between the importance of the cue in signalling focus in the language, and its effectiveness in implying alternatives. Successful communication requires listeners to draw inferences about a speaker’s implicit meaning. This research aims to better understand an important part of these inferencing processes, looking at the linguistic mechanisms listeners are using to generate alternatives.
Table 1. *Focus conditions and probe types used in experiments (underline shows nuclear prominence)*

<table>
<thead>
<tr>
<th>Sentence conditions</th>
<th>canonS Prosodic focus marking only</th>
<th>canonO No focus marking</th>
<th>ScleftO Syntactic focus marking only</th>
<th>ScleftS Syntactic and prosodic focus marking</th>
</tr>
</thead>
<tbody>
<tr>
<td>English example</td>
<td>The wife bought the clock.</td>
<td>The wife bought the clock.</td>
<td>It was the wife who bought the clock.</td>
<td>It was the wife who bought the clock.</td>
</tr>
<tr>
<td>Chinese example</td>
<td>#</td>
<td>妻子购买了钟表。</td>
<td>是(SHI)妻子购买的(DE)钟表。</td>
<td>是(SHI)妻子购买的(DE)钟表。</td>
</tr>
<tr>
<td>Probe types</td>
<td>Identical: wife (identical probes only in the Chinese experiment)</td>
<td>Contrastive alternative: husband</td>
<td>Noncontrastive associate: marriage</td>
<td>Unrelated control: toll</td>
</tr>
</tbody>
</table>

References


The influence of rhythm and animacy on word order in three different age-groups

Isabelle Franz\(^1\), Gerrit Kentner\(^1\), Luisa Bernius\(^2\), Frank Domahs\(^3\)

\(^1\)Max Planck Institute for Empirical Aesthetics Frankfurt, \(^2\)Goethe University Frankfurt, \(^3\)Philipps University Marburg

Prosodic as well as semantic constraints have an effect on children’s and adults’ speech production. For language acquisition, there are many reports indicating a preference for an alternating rhythm in speech, especially for the avoidance of lapses (two or more unstressed syllables in a sequence, Gerken 1996). Furthermore, there are many findings demonstrating a preference for animate referents to be produced before inanimate ones. An influence of this animacy constraint (ANIM) was shown (among others) by Prat Sala and colleagues (2000), and Drenhaus and Féry (2008). Our results show first indications of the constraints’ interaction and influence on the word order of German speaking children (and adults).

We tested 18 children from kindergarten aged three to six years, 13 children from elementary school aged six to ten years, and a control group of 24 young adults. In a picture naming task, participants were instructed to produce coordinated noun phrases (e.g.: ‘dolphin and planet’) without any prespecified order of the conjuncts. As target-items, we used bisyllabic nouns with stress pattern (trochaic, iambic) and animacy (animate, inanimate) as varying factors. Stimuli were diagonally arranged picture pairs with one item placed lower left and the other upper right. We analyzed the sequences the participants chose, yielding either violations of *LAPSE (Rätte und Planät, ‘rat and planet’), or ANIM (Planét und Rätte, ‘planet and rat’), or both (Höse und Delfín, ‘trousers and dolphin’) or none (Delfín und Höse, ‘dolphin and trousers’) to examine the constraints’ influence on sequencing the nouns within a phrase.

Overall, the kindergarten children preferably produced animate items before inanimate ones, showing a significant influence of ANIM on word order \((z = 4.654, p < 0.001)\). The prosodic constraint also showed some impact on the linear order so that *LAPSE constructions were avoided. These results were only significant in the subset where animacy didn’t vary as a factor – that is those item pairs with two animate or two inanimate items \((z = 2.423, p = 0.0154)\).

For the group from elementary school and the adult control group, prosody did not play a significant role for word order, whereas animacy was even more important for ordering the nouns (elementary school: \(z = 4.75, p < 0.0001\); adults: \(z = 4.71, p < 0.0001\)). However, in both groups the effect of animacy was even stronger in the subset where prosody was no varying factor – that is those item pairs with two iambs or two trochees. This suggests that prosody is at least a confounding factor for the power of animacy. Interestingly, spatial order was important for all groups, whereas the younger children preferred to front the upper right picture and the older children the bottom left one. For the adult group, this left-right bias was even stronger, confirming Knudson and colleagues (2014) who found that the left-right bias increases with age.

To conclude, the influence of prosody seems to reduce with age, whereas animacy is a very stable factor for the serialization of conjuncts. Further, the rather weak effect of *LAPSE may be due to shorter naming latencies for trochees (Schiller et al. 2004) and the need in our design to name iambs first when lapses shouldn’t occur so that there could be a counteracting effect of prosody. In any case, our findings can be taken as evidence for the prosodic licensing hypothesis (Demuth 2007), according to which children show a propensity for prosodically well formed structures. In summary, our results suggest the ranking ANIM >> *LAPSE for German speaking children and adults; this ranking corresponds with the one derived from the findings by McDonald and colleagues (1993) for English speaking adults.
Figure 1. Example picture pair (left panel), also presented in reversed order (middle panel), and a filler pair (right panel) in opposed spatial order.
Target Structure picture pair: Klavier und Käfer / Käfer und Klavier (‘piano and beetle’).
Violations depending on word order: *LAPSE: Káférer und Klavíer, ANIM: Klavíer und Káfer
Target Structure filler pair: Mütze und Pirat / Pirat und Mütze (‘hat and pirat’)

References

Focus serves to highlight contrastive information, and this can be expressed through syntax or prosody, depending on the language. In English, a focused word is typically pitch-accented, with increased pitch (F0), duration and intensity [1]. However, previous findings regarding children’s perception and production of focus are mixed [2], suggesting that children’s prosodic skills are protracted [3]. However, the stimuli in these studies differed in the number of syllables and sentence position of the focussed element. This raises questions regarding how and when children learn the acoustic cues to focus.

For adults, we expect adjectives with focus to be higher in F0 and longer in duration than their counterparts without focus [1]. The nouns would manifest in the same way. This would result in an interaction between focus on adjectives vs. nouns. Given the reported protraction of prosodic skills in children [3], we predicted that they might not yet use these acoustic cues to focus in an adult-like manner.

Twenty-four monolingual Australian English (AusE) speaking adults, with a mean age of 20 years (6M, 18F), and 20 monolingual AusE-speaking children with a mean age of 6 years (9M, 11F) participated in the study. The stimuli consisted of 4 utterance-medial adjective + noun phrases, e.g., ‘green ball’. ‘Focus’ was assigned to either the adjective or the noun in 2 experimental conditions: (1) ADJ-FOCUS, (2) NOUN-FOCUS, for a total of 8 stimulus sentences. To ensure the utterance medial position of the focused word, stimuli with the ADJ-FOCUS were embedded in the carrier sentence ‘I have a/an ADJ Noun’, and stimuli with the NOUN-FOCUS were embedded in the carrier sentence ‘I have a/an Adj NOUN now’. The visual stimuli were pseudo-randomized and printed on cards as coloured pictures to be used in a game. The participants then described the picture on each card that differed from that on the experimenter’s card in either the colour or the type of object. Participants’ productions were recorded at 44.1KHz in Audacity. The onset and offset of the adjectives and nouns containing different types of segments were annotated in Praat [4], according to the acoustic landmarks in [5]. F0 in semitones and duration of the adjectives and nouns were extracted for statistical analysis in R [6]. Two linear mixed effects models were fitted to F0 and duration respectively, using lme4 and evaluated using the anova function in LmerTest. Factors included Focus (ADJ vs. NOUN), Word (adjective vs. noun) and Group (Adults vs. Children), with subjects as a random factor.

The F0 model with random intercepts revealed a significant Focus-by-Word interaction (F=94.75, p <.0001), Word-by-Group interaction (F=6.08, p=.014) and Focus-by-Word-by-Group interaction (F=7.58, p <.006). Counter to our prediction, children used F0 to signal focus in an adult-like manner (Figure 1). The Duration model with random intercepts, however, revealed a significant main effect of Group (F=7.314, p=.007) and a significant Focus-by-Word interaction (F=6.796, p =.009). As predicted, children did not use duration in the same way as adults to signal focused nouns (Figure 1). These findings suggest that 6-year-olds are able to use F0 to signal focused nouns, but not duration. This is in agreement with [7], which observed that younger children cannot use accentual lengthening. This raises the question of when children learn to use duration to signal focus/accent.
Figure 1. Pitch (semitone) and durational correlates of focused adjectives vs. nouns in utterance medial adjective+noun sequences in adults vs. children (+/- 2 standard error (SE)).


