Some Questions and Answers on the Prosodic Correlates of Information Structure

Hannes Pirker and **Friedrich Neubarth**

Austrian Research Institute for Artificial Intelligence (ÖFAI)
Schottengasse 3, A-1150 Vienna, Austria
{hannes,friedrich}@oefai.at

Abstract

In this paper a study on the effects of varying information structure and syntactic structure on prosody is presented. For this purpose a corpus of German question-answer pairs was designed and established. The structure and encoding of this corpus is described and some analysis results are presented.

1 Introduction

It is undisputed, that syntactic structure, "information structure" and more specifically the position of "focus" strongly influence prosody ([1]).

Unfortunately, though, there is less agreement on how to define "focus" properly, let alone what its exact effects on prosody are. In addition, the huge number of interdepending influencing factors complicates the analysis of duration in particular ([2], [3]). This makes it difficult to identify and quantify the effects of a single factor. One of the few widely accepted assumptions on information structure is, though, that focus can be controlled by using wh-questions: The constituent corresponding to the wh-word in the question, should - per definitionem - be focused in the answer.

In order to shed more light on the effects of varying information structure and syntactic structure, a corpus of question-answer pairs was constructed. This allows for the manipulation of focus conditions on lexically identical sentences. Thus both the "focus definition"-problem, and the "degrees of freedom"-problem of influencing factors is diminished.

The current work is motivated by two different scenarios: Advantage of its rigid structure is taken within a project, where machine-learning techniques are applied for producing duration-models for speech synthesis. In this context the sub-corpus allows for complementary analysis which help to fine-tune the encoding and selection of factors prior to automatic learning ([4]). The precisely defined information structure, on the other hand, is particularly valuable for improving prosodic models for scenarios which use concept-to-speech generation.

2 Corpus Design

2.1 Textual material

In order to also facilitate the study of the effects of syntactic variation, and to trigger variation in phrasing, the sentences used in our corpus are organised into pairs like (1A) and (1B). All sentences display the same uniform structure namely "**np1 vfin np2 vinf1** *und/and* **np3 vinf2**", with the abbreviations standing for noun phrases (np), finite (vfin) and infinite verbs (vinf) respectively, with the numbers specifying their linear ordering within the sentence. Their assignment to the sentences should become clear from the example (1A/1B) below.

	np1	**vfin**	**np2**	**vinf1**
(1A)	Peter	verspricht	dem Freund	zu verweilen
	Peter	(promises	the friend$_{Dat}$)	to stay$_{Vi}$
	Peter promises (to) the friend to stay...			
(1B)	Peter	verspricht	den Freund	zu entlasten
	Peter	promises	(the friend$_{Acc}$	to relieve$_{Vt}$)
	Peter promises to relieve the friend...			

	np3	**vinf2**
...	und den Dieb	zu bewachen
...	and the thief	to guard
	... and to guard the thief	

10 pairs with this structure form the textual basis of the corpus. The text in these sentences is constructed as follows: All words at position np1 (grammatical subject) are disyllabic proper nouns (e.g. *Peter, Lisa, Heike,...*). At positions np2 and np3 common nouns with varying numbers of syllables are used (e.g. *Freund, Chefin, Beamten, Gastgeberin, Wanderprediger*) and the same words are systematically interchanged between position np2 and np3. For positions vfin, vinf1 and vinf2, each slot has its distinct non-overlapping set of finite (vfin) and infinite (vinf1, vinf2) verbs.

ISBN 1-876346-48-5

2.2 Syntactic variation

It is evident that sentences (1A) and (1B) are almost identical in their spelling, though syntactically and semantically quite different. All 10 A/B pairs in the corpus only differ in the verb chosen for position **vinf1**. In sentences of type A this always is an intransitive verb (e.g. *to stay, to sleep*) and the noun phrase at position **np2** (e.g. *dem Freund*) is thus the indirect object of the finite control verb **vfin**. In its sibling of type B, **vinf1** is exchanged for a *transitive* verb (e.g. *to seek, to catch*) hence **np2** is its direct object. The change of grammatical case of np2 between accusative and dative is reflected in a change at np2's definite article (e.g. *dem* vs. *den*) but leaves the noun itself unchanged. The round brackets in the example's gloss indicate the resulting difference in the internal structure.

2.3 Variation on information structure

The main motivation of the current study was the examination of the prosodic effects of different focus conditions. In our experiment wh-questions are used for imposing the following 10 different focus conditions upon the sentences described above:

- **f_broad** Broad focus on whole sentence ('out of the blue'-utterance):
  What is happening?
  Peter verspricht dem Freund ...
- **f_v2** Focus on the infinitive construction controlled by vinf
  WHAT does Peter promise to do?
- **f_np1** Narrow focus on np1 (i.e. on the grammatical subject)
  WHO promises the friend to stay?
  PETER verspricht dem Freund ...
- **f_np2** Narrow focus on np2 (i.e. on the object of either vfin in type A and vinf1 in type B sentences)
  WHOM does Peter promise to stay?
  Peter verspricht dem FREUND ...
- **f_np2cntr** Contrastive focus on np2
  Does Peter promise the BOSS to stay?
  (No!) Peter verspricht dem FREUND ...
- **f_np2topi** Narrow focus on subject which is placed in position np2 due to topicalisation of the object. This is the only focus condition where the word order is switched: positions np1 and np2 are interchanged.
  WHO promises the friend to stay?
  Dem Freund verspricht PETER zu verweilen ...
  The friend$_{Dat}$ promises PETER$_{Nom}$ to stay ...
- **Variants with added focus particle**: For also examining the additional effects of focus inducing particles, slightly modified versions of the basic sentences were produced, by inserting *sogar/even* between vfin and np2. These *sogar*-versions then were subject to the following 4 focus conditions **f_broad_part**, **f_v2_part**, **f_np1_part**, and **f_np2_part** analogous to the procedure above.

Altogether this design results in an overall number of 250 sentences.[1]

3 Corpus construction

3.1 Recording and annotation

The recordings were performed with a single male speaker of "Standard Austrian German". The questions were presented auditorily and in random order and the speaker was provided with the answers on a separate sheet of paper each. Only the answers were included in the corpus. The speech material was manually labelled for phoneme-boundaries, ToBI-labels for accents and prosodic phrasing, and perceived prominence on a scale from 0 to 4.

3.2 Data selection and normalisation

For the sake of conciseness in this paper we restrict our analysis on the prosodic effects on stressed syllables of the content words in the corpus. Thus for each constituent of the sentence (np1, vfin, np2, etc.) exactly one representative duration value is available.

In order to use only sentence pairs which are lexically completely identical, sentences with focus-type f_np2topi (topikalised) and f_*_part (i.e. containing *sogar*) were excluded. Thus in this paper 100 sentences are analysed: 10 pairs which are subject to the 5 focus conditions f_np1, f_np2, f_np2ctr, f_v2 and f_broad.

The the highly uniform structure of the corpus allows to apply normalisation methods on duration-values prior to statistical analysis, which helps to reduce the problem, that prosody is determined by a number of possibly interacting factors influence on the segmental, syllabic and supra-syllabic level [3]. The normalisation of the durations is performed on the basis of the *mean* duration of those syllables which appear in the same position within identically spelled words in identical sentence position.[2]

For the analysis of the effects on pitch, the extracted F_0 contour was first smoothed and then the *mean* F_0 in the nucleus of stressed syllables was used as a measure.

[1] The full list of sentences together with a detailed explanation of the focus conditions and sound samples can be found at: `http://www.oefai.at/oefai/nlu/speedurcont`

[2] A *normalised syllable duration* (YtpDurN) of "1.2" for a specific stressed syllable (e.g. *'Freund')* in position np2 means this syllable is 20% longer then the average of all token of *'Freund'* appearing in np2 position. The latter condition is added, because some words appear in both np2 and np3.

ISBN 1-876346-48-5

4 Analysis and results

In the confined context of this paper we are trying to present literally a general picture on how the different focus conditions were prosodically realised by the speaker. For this purpose boxplots for both normalised syllable durations (Figure 1) and F_0 (Figure 2) are supplied.[3]

The plots for duration and pitch are organised in parallel. The top three plots give the results for the 3 selected focus conditions: narrow focus on np1 and np2 and broad focus. These are "longitudinal" views. Thus the plots on the right give an impression on the overall pitch contours for the different foci, while the plots on the left show a "durations-contour" accordingly. E.g. in the plots for the narrow foci f_np1 and f_np2 the steep decrease in pitch right after the narrowly focussed constituent can be observed. These plots are enhanced with information on prosodic phrasing: The labels *'IP:'* indicate the percentage of phrase boundaries observed in this position, and this value is also reflected in the height of the line attached to label.

The lower three plots in both figures represent a "cross section" at the positions np_1, np_2, np_3. Here the effects of different foci at a specific position can be compared directly.

These plots indicate, that for the conditions *narrow* and *contrastive* focus on np2 (i.e. f_np2 and f_np2cntr) neither duration nor pitch differ significantly and that they can probably be collapsed into one category.

Another result which also can be quite easily grasped from the plots is the distinguished behaviour of sentence initial narrow focus f_np1: This is the only condition which displays highly significant effects in both duration and pitch and these are not confined to the locus of focus np1 but also is present in position np2. This distinguished effect of f_np1 is not spread as far as np3, the penultimate constituent in the sentence which is never focussed. In this position there only is a significant difference in pitch between preceding narrow (f_np1, f_np2*) and non-narrow focus (f_broad, f_v2).

A result that is not fully compatible with prior expectations is the prosodic phrasing displayed in the examples: there is much less variation observed than would be expected, and the number of boundaries is exceedingly high. Note that in theory a boundary between vfin and np2 should appear in at most 50% of the sentences, because it should not appear in type A sentences at all. This constraint is clearly violated. This is partly due to "cognitive overload" for the speaker when concentrating on the production of narrow foci and of list-effects that can be observed due to the uniformity of stimuli. But also the quality of the labelling has to be assessed: currently it is not distinguishing between major and minor phrases, and many of the marked boundaries are quite weak. An improved labelling scheme, that takes the obvious differences in boundary-strength into account should thus be employed for future investigations.

[3] All statistics and graphics were produced with the invaluable open source software **R** (`http://www.r-project.org`)
In boxplots a box's middle line corresponds to the *median* (not the *mean*), upper and lower box borders designate *quantiles*, and the lines extending from it indicate the *min* and *max* (with possible outliers marked as dots). When notches are used, the non overlapping of notches indicates significant differences.

5 Conclusion

We presented a highly structured corpus of question-answer-pairs which allows for the systematic analysis of the prosodic realisation of different foci. Apart from this, the corpus seems also to be quite promising for applying detailed investigation on domain-edge and domain-span effects along the lines of [3] in the future.

Acknowledgments

We express our gratitude to our former colleagues Kai Alter, who designed the corpus and conducted its recordings and Elli Rieder who participated in the tedious labelling. The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Education, Science and Culture. The production of the corpus was sponsored by the *Fonds zur Förderung der wissenschaftlichen Forschung (FWF)*, Grant No.P13224. The work reported in this paper is supported by the EC Project NECA IST-2000-28580. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

References

[1] Rump H., Collier R.: Focus Conditions and the Prominence of Pitch-Accented Syllables, Language and Speech, 39 (1),pp.1-17, 1996.

[2] Campbell N.W.: Syllable-based segmental duration, in Bailly G., Benoit C. (eds.), *Talking Machines*, North-Holland, Amsterdam/New York, pp.211-224, 1992.

[3] White L.: English speech timing: a domain and locus approach, University of Edinburgh, UK, 2002.

[4] Neubarth F., Pirker H., Trost H.: Learning duration, in Busemann S.(ed.), Konvens 2002, DFKI, Saarbrücken, Germany, pp.123-130, 2002.

ISBN 1-876346-48-5

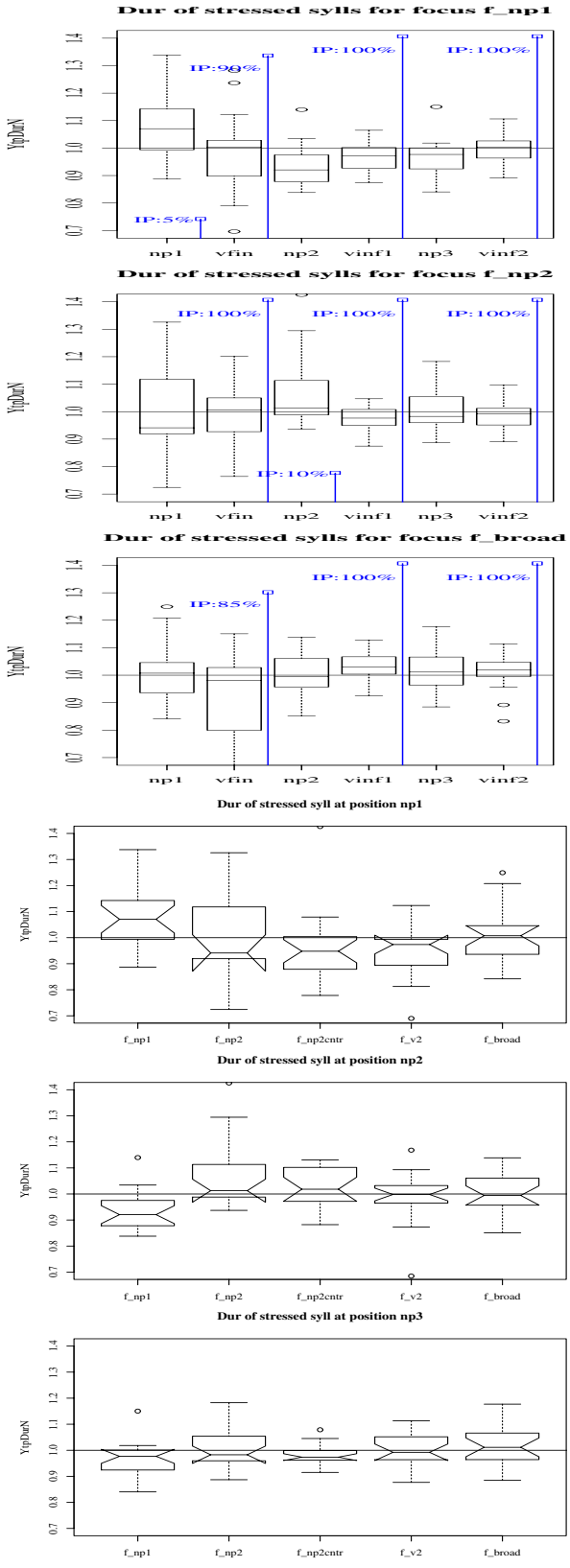

Figure 1: Normalised durations of stressed syllables plotted in 'longitudinal' (top 3) and 'cross-section' manner (bottom 3 plots). N=100 sentences. The vertical lines and the numbers labelled 'IP:' point out the frequency of prosodic phrase boundaries .

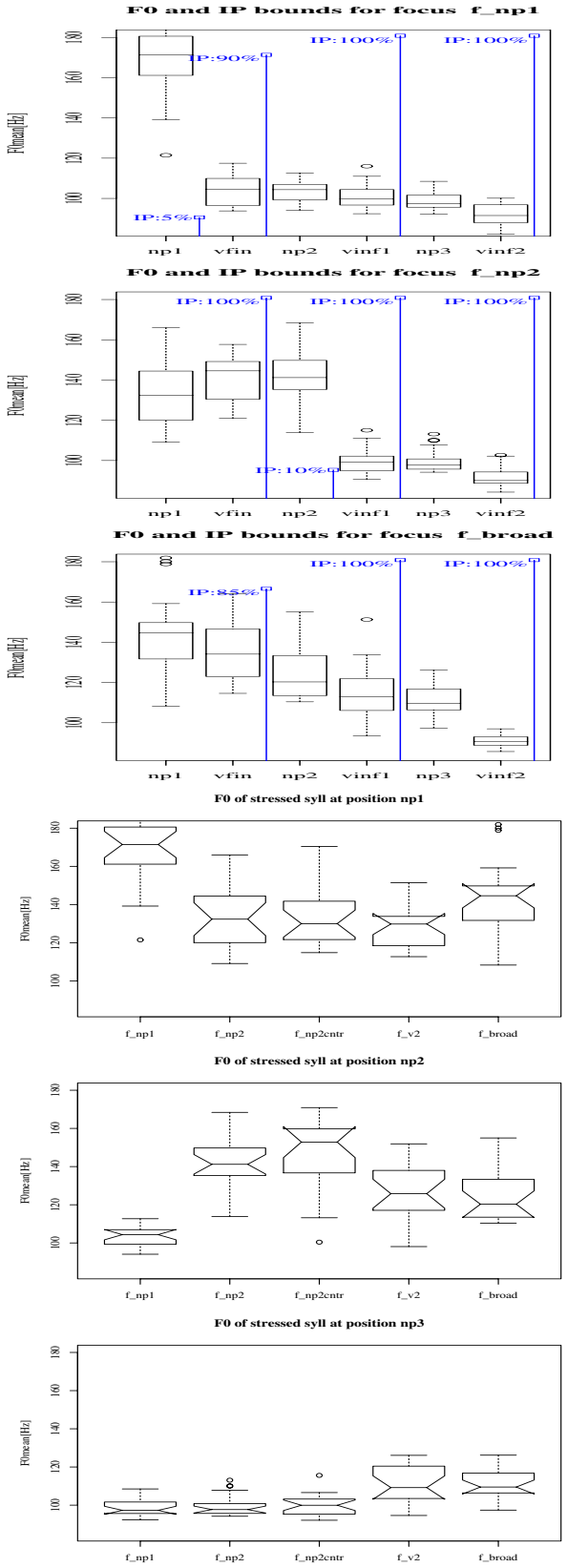

Figure 2: Mean F_0 within the stressed syllables' nuclei plotted in a 'longitudinal' (top 3) and a 'cross-section' manner (bottom 3).N=100 sentences.

ISBN 1-876346-48-5