In Reich, Ingo *et al*. (eds.), *Proceedings of Sinn & Bedeutung 15*,
pp. 193–208. Universaar – Saarland University Press: Saarbrücken, Germany, 2011.

Focus Marking via Gestures*

Cornelia Ebert
University of Stuttgart
cornelia.ebert@
ling.uni-stuttgart.de

Stefan Evert
University of Osnabrück
stefan.evert@uos.de

Katharina Wilmes
University of Amsterdam
K.Wilmes@
student.uva.nl

Abstract. This paper contributes to the recent investigations of speech-accompanying gestures under a formal semantic view. We show that gestures can serve to disambiguate a sentence with respect to its possible focus domains. We provide a statistical evaluation of data gained from a corpus annotated with gestures and information structure. The language under investigation is German. We argue that a sentence that, in isolation, is ambiguous concerning the extension of its focus domain is disambiguated via speech-accompanying gestures. Gesture thus is a means to mark information structure next to intonation and word order.

1 Introduction

It is widely known that gestures are temporally aligned with the speech signal, in particular it has often been claimed that the *stroke*, i.e. the main part of a gesture where the actual gesture movement takes place, falls together with the main accent of the gesture-accompanying sentence (McNeill 1992 among many others). The relationship of complete gestures or *gesture phrases* and foci, however, has not been investigated systematically yet. We want to fill this gap by showing that the possible focus projection of a focus exponent is restricted by the point of time at which a speech-accompanying gesture starts. Gesture thus serves as a means to mark focus domains. Consider the following example for illustration (the main accent is indicated by capital letters):

(1) I ate baNAnas.

* First and foremost, we would like to thank Hannes Rieser and Florian Hahn for giving us access to the gesturally annotated SAGA-corpus of the University of Bielefeld. This work would not have been possible without the possibility to access and make use of the accurate and fine-grained gestural annotations of the SAGA-corpus. We would also like to thank Hannes Rieser and Florian Hahn for their constant help with technical and other questions of all sorts as well as for numerous valuable discussions about gestures and information structure.

The sentence in (1) with the given intonation pattern can be read as an answer to the two questions in (2), each inducing a different focus-background structure.

(2) a. What did you do?
b. What did you eat?

(2a) is a VP-focus invoking question, while (2b) requires narrow focus on the direct complement. Following (2a), (1) allows for the focus pattern in (3a); if (1) follows (2b) on the other hand, the focus pattern is the one of (3b).

(3) a. I [ate baNAnas]$_F$.
b. I ate [baNAnas]$_F$.

In the following we will defend the hypothesis in (4).

(4) Hypothesis (Focus-gesture alignment):
How far a focus projects is determined by the onset of the accompanying gesture (if one exists).

In other words, the onset of a speech-accompanying gesture indicates the left border of the focus phrase (independent of the type of gesture – be it a beat, a deictic or an iconic gesture or any other kind of gesture). A speech-accompanying gesture can thus serve to disambiguate an information-structural ambiguity in a sentence towards a certain focus-background pattern. Simplifying matters for now, we expect the patterns in (5). ($|_G$ marks the hypothesized onset of the speech-accompanying gesture.)

(5) a. I $|_G$[ate baNAnas]$_F$.
b. I ate $|_G$[baNAnas]$_F$.

Although (1) is ambiguous with respect to the underlying information structure, $|_G$ disambiguates the sentence towards one of the focus-background patterns in (3).

In order to test the hypothesis in (4), we looked at the temporal occurrences of gestures and foci. We therefore annotated the multimodal Bielefeld Speech-And-Gesture-Alignment (SAGA) corpus with focus features – in addition to the existing gestural annotation – and marked the nuclear accents of certain intonation units. A subsequent statistical analysis confirmed our hypothesis that the onsets of focus and gesture align indeed – with a systematic shift, however: on average gestures start about 0.3 seconds earlier than the corresponding focus phrases. That is, there is a certain time lag between the onset of a gesture and its associated focus.

In this paper, we mostly present material that has also been discussed in Wilmes (2009). We re-evaluate some of the results of Wilmes (2009) and further elaborate on various aspects. The remainder of this paper is structured as follows: Section 2 sets the stage and discusses the relevant findings from the gesture literature that will be needed in the remainder of the paper. Section 3 presents the methodology underlying our investigations. Here, we explain what the data set that our study is based on looks like, how we annotated these data and how we finally investigated the temporal interdependence of gestures and foci. Section 4 then presents the results of a statistical investigation of the temporal occurrences of gestures and foci. In section 5, we evaluate and discuss these results. Section 6 discusses some controversial issues and loose ends. And finally, section 7 concludes the paper.

2 Speech-Accompanying Gestures

It is a widely held view that gesture is a distinct mode of expression and that the study of gestures can tell us more about language than one might think at first sight (see e.g. Kendon 1972, 1980 and Loehr 2004 and references therein). We subscribe to this view and we will argue in particular that for a comprehensive view of focus phenomena it is inevitable to take speech-accompanying gestures into account.

To set the stage, we will have a look at some important findings concerning the interpretation of speech-accompanying gestures. First of all, one has to define what a gesture phrase is, i.e. where it starts and where it ends. In order to determine which movements can be considered to contribute to a particular gesture, Kendon (1972, 1980) identified a certain structure that can be found for gestures quite generally. The smallest unit of a gesture is its main element, i.e. the minimally required element for being reckoned as a proper gesture: the *stroke*. The stroke can be identified with the strongest movement within the gesture. A stroke is usually preceded by a *preparation phase* and followed by a *retraction phase*, for the hands must be brought into an appropriate position for the stroke to be executed and back into the resting position. Taken together, these three phases constitute the gesture phrase[1]. Preparation and retraction are optional, so a gesture phrase may consist of nothing but a stroke. Between preparation and stroke and stroke and

[1] This notion of a gesture phrase cannot be applied to all kinds of gestures. So-called beats are only biphasal, i.e. they consist of two movement phases, constituting a repeated movement pattern, like up and down or in and out.

retraction *holds* may occur, which are termed *pre-* or *postholds*, respectively. These are considered to enhance timing between speech and gesture (cf. McNeill 1992, Lascarides and Stone 2009).

Importantly, it has been argued that gesture and speech can work together to convey one single thought (McNeill 1992, Kendon 1980) and hence that the semantic content of speech-accompanying gestures is intertwined with the semantic content of the speech signal. What is especially important for our purposes is that speech-accompanying gestures are known to be temporally aligned with the speech signal. It has been argued that speech and gesture synchronise in that the stroke of the gesture falls together with the main accent of the gesture-accompanying utterance (see among others: Pittenger, Hockett, & Daheny 1960; Kendon 1980; McNeill 1992; Loehr 2004; Jannedy & Mendoza-Denton 2005). The general claim is that the stroke occurs just before or at the same time as (but not later than) the nuclear accent. Although there are very few empirical studies that back this claim (see Loehr 2004 for a recent study), this is a fairly established finding in gesture theory.

What has been far less investigated is the interaction of entire gesture phrases and speech. In the literature one can find only a few hints and claims concerning their interdependence and there seems to be no general agreement. Kendon (1972: 184) suggests that gesture phrases align with so-called '*tone units*' (i.e. '*the smallest grouping of syllables over which a completed intonation tune occurs*', cf. Loehr 2004). Loehr (2004) on the other hand argues that gesture phrases and '*intermediate phrases*' in the sense of Pierrehumbert (1980) align. We want to add to this list and argue that it is actually focus phrases that gesture phrases align with. Hence, while Loehr (2004) and Kendon (1972) argue that the temporal occurrence of gesture phrases is mainly triggered by intonational aspects, we think that gesture phrases rather synchronise with focus phrases, which means that their temporal appearance is determined by information structure. While there is, of course, a clear connection between intonation and focus, we still believe that the alleged interdependence between gesture phrases and whichever kind of intonationally motivated category is – at best – an epiphenomenon of the gesture-focus alignment for which we argue.

3 Methodology

To verify our hypothesis in (4) that (the onsets of) gesture phrases align with (the left border of) focus domains, we investigated the temporal interdependence of gesture phrases and focus domains. In addition, we also

looked at the timing of stroke and nuclear accent. Our study is one of the very few empirical studies about the interplay between gesture and intonation; to the best of our knowledge, it is the first empirical study of the interplay between gesture and focus. We analysed a 20-minute video sequence with 275 gestures, which makes this study the most extensive empirical study on gesture and speech (cf. Loehr 2004: Condon & Ogston 1966: 5 sec; Kendon 1972: 90 sec; McClave 1991: 125 gestures; Loehr 2004: 164 sec and 147 gestures).

3.1 Data

For our study, we worked with one sequence of the Bielefeld SAGA-corpus (Lücking et al. 2010), which is a multimodal corpus (video and audio) that collects dialogues from an experiment where one subject (the *router*) gives directions to another subject (the *follower*) for navigation through a dynamic virtual world (see Lücking et. al 2010 for details). While talking, the movements of the subjects' hands were recorded by sensors attached to the hands and fingers. Three video cameras recorded the scene from different angles. Sound was also recorded.

From this corpus we selected a 20-minute sequence with two male participants. Gestures were already annotated, including gesture type (e.g. iconic or deictic) and duration of gesture phases (i.e. preparation, stroke, holds and retraction).

3.2 Annotation

For our purposes, it was necessary to add information-structural annotation (accent and focus) to the existing gestural annotation of the selected video. Our annotation was entirely based on the audio material, which had already been transcribed (but not annotated with parts of speech or other morpho-syntactic information). The information-structural annotation was carried out without reference to the video and its gesture annotations in order to exclude a possible bias. We annotated nuclear accents and distinguished two types of foci: *new-information* and *contrastive*. All annotations were based on the recommendations of Dipper et al. (2007) (in particular Chapters *Phonology and Intonation* (Féry et al. 2007) and *Information Structure* (Götze et al. 2007)). We treated as *new-information focus* those cases where information is provided which is new and/or carrying the discourse forward. Here, we predominantly found rather broad focus domains: whole sentences (all-focus sentences), e.g. if these sentences were text-initial or answers to polar questions, and VP-foci. However, our data also contain narrow foci such as

DP- or AdjP-foci. An expression was tagged as *contrastive focus* if it overtly contrasted with other elements in nearby utterances.
We kept track of all pitch accents in the data, i.e. the points of highest or sometimes lowest pitch that make syllables intonationally salient (*X** in the ToBI framework[2]) and filtered out the nuclear pitch accents among them. There was always one unique nuclear accent for each new-information focus domain. For reasons of space, we cannot go into the details of the annotation procedure and refer to (Wilmes 2009: 26-31) for further information.

3.3 Data Extraction

To verify hypothesis (4), i.e. to show that gesture phrases and focus phrases align in fact, we investigated the temporal interdependence of focus phases (*FocPs*) and gesture phrases (*GPs*). This left us with the following task:

(6) Verification Task (Focus-gesture alignment):
For each gesture phrase, find the corresponding focus phrase and compare the temporal position of the two.

For each gesture, we had a look at the associated speech (not the other way round).[3] Making use of the result from the literature that nuclear accents and strokes align, we associated a gesture phrase with a focus phrase if the nuclear accent of the focus phrase overlapped with the gesture phrase's stroke (see *Figure 1* for an example). In the few cases where there was no main accent coinciding with the gestural stroke, we considered a focus phrase overlapping with at least the stroke phase to be associated with the gesture, unless the overlap was very small and a close investigation of the gesture-focus pair made an association implausible (because there was another focus that was more likely to associate with the gesture). This was the case for only two gestures. Moreover, there were eight cases of strokes that did not overlap with any focus. In one case, an entire gesture did not coincide with any focus at all and for seven gestures, though they overlapped with a focus in some parts, it was not the stroke that overlapped with the focus. We excluded these ten gestures and strokes from our statistical evaluation.

Figure 1 illustrates an example that shows how gesture time and focus time can be compared. Time differences are assessed by subtracting focus

---

[2] TOBI stands for Tone and Break Indices. The system is based on work by Pierrehumbert (1980). In our study, we did not distinguish between different kinds of pitch accents like *high* (*H**), *low* (*L**) or *rising* (*L+H**).
[3] Thus, if there is no gesture there is also no need to identify a focus to verify our hypothesis. However, in most cases we found a one-to-one mapping of focus phrase and gesture phrase.

times from gesture times (e.g. start difference = gesture start – focus start). The corresponding sentence from the corpus is given in (7):

(7) Ja, also die Busfahrt, die hat äh fünf Stationen, die auf jeden
Yes so the bus tour RP has eh five stops that on every
Fall angefahren werden müssen.
case approached will must
'Yes, so on the bus tour there are five stops that have to be approached in any case.'

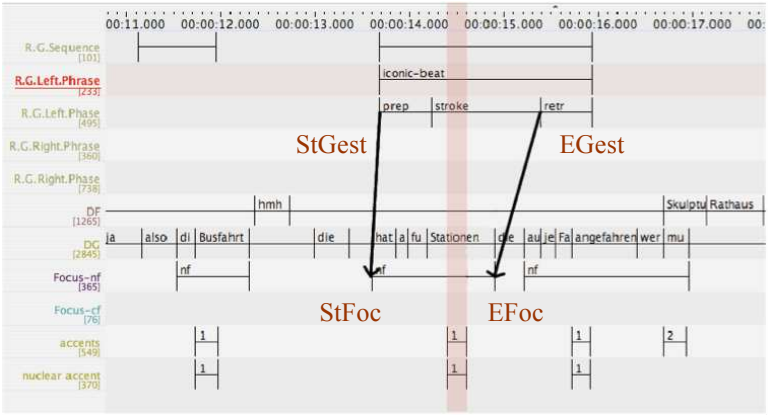

Figure 1*: Comparison of focus and gesture times*

The onset time of the focus phrase (*StFoc*) is subtracted from the onset time of the associated gesture phrase (*StGest*), i.e. the onset of the preparation phase (or the stroke if there is none). The time when the focus phrase ends (*EFoc*) is subtracted from the time when the stroke ends (representing the end of the gesture phrase, hence *EGest*). We treat the end of the stroke and not the end of the retraction phase as the end of a gesture for two reasons: First, according to McNeill (1992: 29) the retraction phase is 'semantically neutral' and second, Loehr (2004) discusses the possibility to disregard retractions and post-holds in his statistical evaluation as well, because they seem to have a different status as the other phases of a gesture phrase.[4]

[4] Cf. Loehr (2004: 117): 'Typically, an entire g-phrase [CE/SE/KW: gesture phrase] aligned with an intermediate phrase. Occasionally, however, it was clear that a g-phrase aligned with an intermediate phrase only when disregarding post-stroke holds, [or] retractions [...] within the g-phrase. These internal components are included within g-phrases by definition, following

As a base for comparison, we also studied the temporal occurrences of nuclear accents (*NAcc*) and strokes in order to verify the by now well-established claim from the literature that nuclear accents and strokes align (cf. section 2). For each stroke, we considered a nuclear accent that overlapped with the stroke as associated with the stroke. If there was no such accent, we took the nearest nuclear accent. Time differences were again calculated by subtracting accent time from stroke time (e.g. start difference = stroke start – accent start).

4 Results

In the following we present our results on the hypothesized gesture-focus alignment and our reassessment of the question whether stroke and main accent align, as has been claimed in the literature. Statistical analysis was carried out with the R environment for statistical computing (R Development Core Team 2005).

4.1 Alignment of Main Accent and Stroke

In total, we analysed 275 stroke-accent pairs. In the majority of cases (209 pairs) the stroke began earlier than the main accent (versus 66 pairs where accent began earlier). Similarly the stroke ended later than the main accent for 183 pairs (versus 92 pairs where the accent ended later). In 124 cases, the stroke encompassed the main accent, in 100 cases stroke and main accent overlapped in some other way, and in 51 cases they did not overlap at all.

Figure 2 shows a histogram for the time difference between the onsets of nuclear accents and the corresponding strokes.

As can be seen, the distribution is approximately Gaussian (the solid line shows the empirical distribution, the dashed line a Gaussian approximation). On average, the stroke starts 0.36s earlier than the corresponding nuclear accent. The standard deviation is about 0.55s. We interpret this as a tendency for gestures to precede the corresponding accent (though there are a

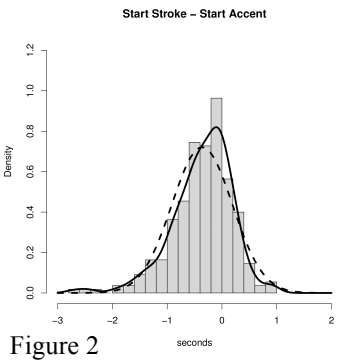

Figure 2

Kendon's hierarchical packaging. However, there may be some different quality about these post-stroke components. Occurring after the heart of the gesture, they may have a less important status in terms of timing with speech.'

considerable number of cases where the gesture starts later).

For comparison, the offset differences have a mean of 0.53s (i.e. stroke usually ends later than the accent) and a standard deviation of 1.25s (*Figure 3*). It is obvious that the onsets align much better than the offsets: their standard deviation is considerably smaller. On the whole, we take our results to show that there is indeed an alignment between the beginning of the stroke and the beginning of the main accent, as claimed in the literature.

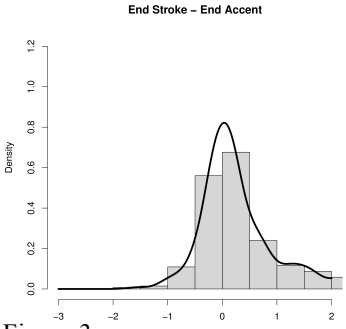

Figure 3

4.2 Alignment of Focus and Gesture

Having obtained experimental confirmation for the alignment of nuclear accents and strokes, we now turn to our hypothesis that gesture phrases and focus phrases are also synchronised. We found that contrastive foci and new-information foci behave somewhat differently with respect to their accompanying gestures, so we evaluated the two types of foci separately. We analysed 260 new-information focus–gesture pairs and 56 contrastive focus-gesture pairs. As pointed out above in Section 3.3, ten gestures were excluded from the analysis because no focus could be associated with them.

4.2.1 New-Information Focus and Gesture

Figure 4 shows the distribution of the onset differences of gesture and new-information focus (we refer to *new-information focus* simply as *focus* in the following), which corresponds almost perfectly to a Gaussian distribution. With 0.41s, the standard deviation is rather small. Again we find a systematic shift: gestures start on average about 0.31s earlier than foci, and there are only few cases where focus precedes gesture. While there is thus a certain time lag, most gesture-focus pairs are within less than one second of each other and can be considered to be aligned. A one-sample t-test shows that the time lag effect is genuine ($t=12.41$, $df=259$, $p < .001$; H_0: mean time lag = 0). The corresponding

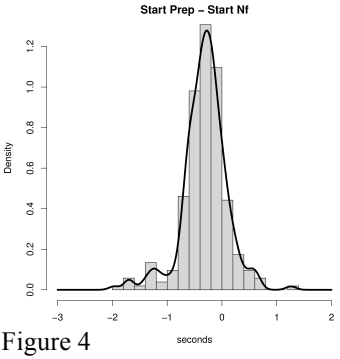

Figure 4

95%-confidence interval places the true mean time lag between gesture and focus in the range from 0.264s to 0.363s.

We consider these results as a confirmation of our hypothesis (4) that gestures and foci align in their onsets.

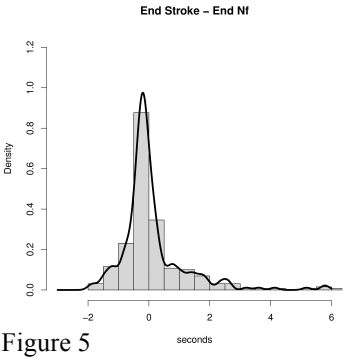

Figure 5

For the offsets, the situation is not as clear. *Figure 5* shows the distribution of the time differences between the end of a gesture (i.e. the end of the stroke) and the end of the corresponding new-information focus. With a mean of –0.15s, there is no evidence for a systematic shift. The standard deviation of 1.24s, however, is comparatively huge, and some gestures end several seconds after the corresponding focus phrase. On the basis of our data, offsets of gestures and foci thus do not seem to synchronise.

4.2.2 Contrastive Focus and Gesture

For contrastive foci and the accompanying gestures, the alignment was not as neat as for the new-information foci. *Figure 6* shows a histogram of the onset differences between gestures and contrastive foci. With 0.70s the standard deviation is rather high. The mean is –0.77s, so gestures have a clear tendency to start earlier than the corresponding foci. We interpret these data to show that there is no tight alignment between the onsets of contrastive foci and those of the associated gestures. We also tested whether contrastive foci align with the stroke rather than the entire gesture. The histogram for the onset differences of contrastive foci and strokes is given in *Figure 7*.

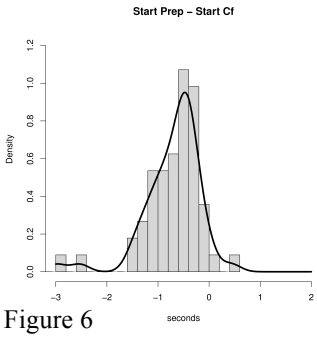

Figure 6

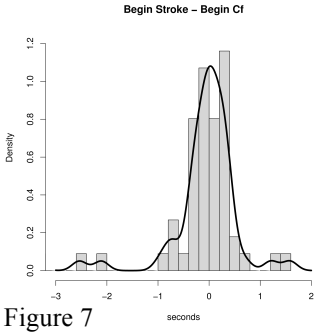

Figure 7

Again, the standard deviation is quite large (0.75s), but in this case there is no evidence of a systematic shift (mean lag = –0.11s). With such high variability, it is impossible to interpret these results as evidence for an alignment of contrastive foci and strokes.

To conclude, we have not found any focus-gesture or focus-stroke alignment effects for contrastive foci. One has to keep in mind, though, that our data set of contrastive foci is rather small. We therefore leave a detailed investigation of contrastive foci and their accompanying gestures for future research, which will need to build on larger amounts of empirical data in order to draw any reliable conclusions.

5 Discussion

In the following we will briefly discuss and evaluate the results that we presented in Section 4. Since our data set for contrastive foci is too small to draw reliable conclusions, we limit our discussion to the comparison of new-information foci and gestures as well as nuclear accents and strokes.

5.1 Shift Effect

As indicated above, our observation that strokes usually start 0.36s earlier than the corresponding nuclear accents is entirely in line with the claims from the literature, where it has been noted that a stroke usually coincides with or starts earlier than its corresponding nuclear accent, but in general does not start later than the accent (Kendon 1980, McNeill 1992). We found the same type of shift for gesture phrases and focus phrases, too. Gestures usually start 0.31s earlier than the corresponding focus domains. We believe that this significant time shift may have its roots in the fact that it allows the hearer to draw attention to the upcoming focus phrase, as its occurrence is made predictable by the preceding gesture. Moreover, it is plausible to assume that gesture production is faster than speech production and that the time lag between the onsets of speech and gesture is due to this difference in generation complexity (cf. also Loehr 2004: 29).

5.2 Alignment

We interpret our results above as support for hypothesis (4), i.e. they show that gesture phrases and (new-information) foci align (with a certain time lag). We still need to clarify what exactly counts as '*alignment*', though. Our main arguments supporting the gesture-focus alignment hypothesis are as follows. First and foremost, we take the stroke-accent alignment, which is a well-established effect from the literature, as a point of reference. The onset differences between nuclear accents and strokes have a mean of –0.36s and a standard deviation of 0.55s. Our results show a considerably better gesture-focus alignment, with a similar shift of –0.31s and smaller standard deviation (0.41s). Compare the corresponding histograms in *Figures 2* and *4*: the better alignment of gesture and focus is immediately obvious.

There is a second argument to support the interpretation of our results in favour of hypothesis (4). As to our knowledge, there is one empirical survey that our study can directly be compared with (Loehr 2004). When interpreting his results, Loehr (2004) was confronted with the same problem, i.e. to define what exactly can be considered as an alignment. He found that the so-called *apex* (the peak of a stroke) and the main accent coincide with a standard deviation of 0.27s (and without any significant shift). He interpreted this as showing that there is a tight alignment of apex and nuclear accent. Furthermore, he also suggested that there is an interdependence of Pierrehumbert's (1980) intermediate phrases and gesture phrases. Similar to our results for gesture phrases and focus phrases, he found that gesture phrases usually start before the corresponding intermediate phrases. The standard deviation for the onset differences between intermediate phrases and gesture phrases was 0.55s. As Loehr (2004) interpreted his results as evidence for a genuine alignment, we think that our study (with standard deviation of only 0.41s) can safely be interpreted to show an alignment of gesture and focus, too.

We did not find evidence for a corresponding alignment of the offsets of gestures and focus phrases. With 1.24s, the standard deviation was very large (recall that the end of a gesture is defined as the end of the stroke). Looking at the histogram in *Figure 5*, however, it seems that for some gestures there is a good alignment (the main peak of the histogram), while for others the stroke is held much longer (the long right-hand tail of the histogram). This suggests that there may be two different types of gestures – one that aligns well with the focus of the accompanying speech signal and another type that does not. We have not investigated this possibility in depth yet, but it would be worthwhile for future research to examine whether there are certain types of gestures (e.g. beats, deictics and iconic gestures) whose purpose it is to structure information and which thus align better with the speech signal than others (e.g. discourse gestures) that might serve a different purpose.

Finally, let us briefly point out once again that we did not reach a conclusion with respect to contrastive foci. We would need more data in order to see how they relate to the accompanying gestures (see Section 4.2.2 for a discussion) and we hope that future research will shed light on this question.

6 Further Issues

Some issues are still open for discussion and call for further research. In the following, we address some of these topics. In particular, we want to point

out that the alignment of focus phrases and gesture phrases is 'real' and not merely an epiphenomenon of some underlying alignment effect of a different nature.

6.1 A Qualitative Argument

It has been proposed in the literature that gesture phrases align with 'tone groups' (Kendon 1972) or 'intermediate phrases' (Loehr 2004), cf. section 2. We have now added another suggestion: gestures align with focus phrases. However, it is possible that none of these claims are true, and that gestures are simply synchronised with certain syntactic categories, e.g. entire sentences or VPs. As our corpus predominantly consists of all-foci sentences and VP-foci, this possibility cannot be excluded without further inspection. Unfortunately, the SAGA corpus is not syntactically annotated, so a quantitative evaluation of how well different kinds of syntactic categories align with gestures cannot easily be carried out without time-consuming manual work. However, we attempted a qualitative assessment of this question. We took a closer look at narrow foci and foci that begin a considerable time later than the corresponding utterance and checked how well they align with an accompanying gesture. We found that if a focus does not begin at the start of the utterance, the corresponding gesture also begins at some later point in nearly all cases. In (8) we give some examples in point:

(8) a. genau äh also [e|$_G$rst Kreisverkehr]$_F$
exactly eh so first roundabout
'exactly, eh, first the roundabout'

b. die haben beide |$_G$[dieselben Türen und dieselben Fenster]$_F$
they have both the same doors and the same windows
'they have both the same doors and the same windows'

c. rechts von dieser Kap|$_G$elle [ist ein großer Laubbaum]$_F$
right of this chapel is a big broadleaf tree
'to the right of this chapel there is a big broadleaf tree'

In all three example cases, the gesture starts near the start of the focus phrase and not at the beginning of the utterance. The gesture phrase thus seems to be aligned with the focus phrase and not with the entire utterance. Furthermore, we found no evidence for a general alignment of gesture phrases with any syntactic categories such as sentences or VPs (see Wilmes 2009 for details).

6.2 A Quantitative Argument

Here, we attempt to show that the alignment of gesture phrase and focus phrase cannot be a secondary effect of the well-known stroke-accent

alignment and the fact that the initial part of the focus phrase (up to the main accent) and the preparation phase have similar lengths. Note that the time difference Δt_F between onset of gesture and focus phrase is the sum of the time difference Δt_A between onset of nuclear accent and stroke and the length difference Δl between preparation phase of the gesture and focus phrase up to the main accent. Assuming that Δt_A and Δl are independent alignment effects, we would expect the standard deviation of the resulting gesture-focus alignment Δt_F to be greater than the standard deviations of Δt_A and Δl. This is not the case: the standard deviation of Δt_F was only 0.41s in our study, whereas the expected standard deviation would be 0.82s (see Wilmes 2009 for details on this calculation). Moreover, we would then expect a strong correlation between the time differences Δt_F and Δt_A as well as Δt_F and Δl, while Δt_A and Δl themselves should be independent or weakly correlated. Our data show an opposite effect: there is only a weak correlation between Δt_F and Δt_A (Pearson's $r \leq 0.219$), but a very strong correlation between Δt_A and Δl (Pearson's $r = 0.759$). From these results and the pairwise correlation plots (omitted for lack of space), we conclude that the length differences arise from two independent alignment effects for stroke and main accent, and for gesture and focus phrase.

7 Conclusion

In our study, we were able to verify claims from the literature that gestural strokes and nuclear accents align (albeit with a systematic shift). We also found a clear, but shifted alignment for the onsets of gesture phrases and (new-information) foci. We interpret these results to show that gestures are a means of marking information structure next to intonational and syntactic means, i.e. speech-accompanying gestures can indicate focus domains.

Furthermore, we were able to show that gestures can serve to disambiguate. A sentence that is information-structurally ambiguous in isolation can be disambiguated by its accompanying gestures. This is yet another observation suggesting that ambiguity might be less of a problem for natural language than was originally thought. While many sentences (e.g. simple SVO sentences with two quantifiers) that seem ambiguous at first sight are disambiguated via intonation in natural speech, we showed that sentences that seem ambiguous even when intonation is taken into account are in fact disambiguated by accompanying gestures.

We hence support the view of Lascarides and Stone (2009) that a formal semantic model should represent not only the usual semantics of linguistic

expressions, but also take care of the semantic contribution of their accompanying gestures.

References

Condon, William & W. Ogston. 1966. Soundfilm analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease* 143. 338–347.

Dipper, Stefanie, Michael Götze & Stavros Skopeteas. 2007. Information structure in cross-linguistic corpora: Annotation guidelines for phonology, morphology, syntax, semantics, and information structure. ISIS 7. 147–187.

Jannedy, Stefanie & Norma Mendoza-Denton. 2005. Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure*, 3. 199–244.

Kendon, Adam. 1972. Some relationships between body motion and speech: An analysis of an example. In A. Siegman & B. Pope (Eds.), *Studies in dyadic communication*. New York: Pergamon Press.

Kendon, Adam. 1980. Gesticulation and speech: Two aspects of the process of utterance. In Mary Ritchie Key (Ed.), *The Relationship of Verbal and Nonverbal Communication*. The Hague: Mouton.

Lascarides, Alex & Matthew Stone. 2009. A formal semantic analysis of gesture. *Journal of Semantics*. 1–57.

Loehr, Daniel. 2004. *Gesture and Intonation*. PhD thesis, Georgetown University, Washington, DC.

Lücking, Andy, Kirsten Bergman, Florian Hahn, Stefan Kopp & Hannes Rieser. 2010. *The Bielefeld Speech and Gesture Alignment Corpus (SaGA).* In M. Kipp, J.-C. Martin, P. Paggio & D. Heylen (eds.), LREC 2010 Workshop: Multimodal Corpora–Advances in Capturing, Coding and Analyzing Multimodality.

McClave, Evelyn. 1991. *Intonation and Gesture*. Doctoral Dissertation, Georgetown University, Washington DC.

McNeill, David. 1992. *Hand and Mind: What gestures reveal about thought.* The University of Chicago Press, Chicago and London.

Pierrehumbert, Janet B. 1980. *The Phonology and Phonetics of English Intonation*. PhD dissertation, MIT. [IULC edition, 1987].

Pittenger, R., Hockett, C. & Danehy, D. 1960. *The first five minutes: A sample of microscopic interview analysis*. Ithaca, NY: Paul Martineau.

R Development Core Team. 2005. R: A language and environment for statistical computing, reference index version 2.x.x. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Wilmes, Katharina A. 2009. *Focus marking by speech-accompanying gestures*. Bachelor Thesis. University of Osnabrück. Germany.