

External Examiner's Report on the Dissertation of Jakub Jehlička (JJ) "Gesture and Eventuality A Crosslinguistic Study"
Submitted in 2021 at the Department of Linguistics

I. Brief summary of the dissertation

The aim of this dissertation is to examine the relationship between gesture and the linguistic expression of "eventuality". One type of gesture is in focus, namely "beats" that can be "ended" or "continuous". The linguistic expression of "eventuality" is mainly focused on grammatical aspect and related concepts (among them Vendlerian verb classes and "directedness"). Two studies are included. The first study compares an existing multimodal (video) corpus of English faculty-student meetings with a comparable Czech corpus collected by JJ. The second study is an online experiment in which native speakers of Czech viewed videos of ended vs. continuous gestures and selected the best match for the video from among two Czech sentences that differ only in the aspect of the verb (perfective vs. imperfective). Visualizations and statistical models are used in the analysis of both studies. Aspect (along with related concepts) is associated with ended vs. continuous gestures in both studies.

II. Brief overall evaluation of the dissertation

This is a fine dissertation and a real contribution to the field of multimodal linguistics. The specific focus of the dissertation on beat gestures and aspect in Czech and English is original and valuable. JJ clearly demonstrates considerable erudition, especially in terms of theory and relevant studies of gesture and sign language. The inclusion of Peircean semiotics is impressive and refreshing, as it is often overlooked in cognitive linguistics. JJ's ability to interpret and synthesize previous scholarship is quite strong. JJ writes well. JJ is very clever with experimental design and with various types of hardware and software.

The length of the introduction and literature review (108pp., plus several pages of additional literature review in Chapter 6) is excessive in relation to the size of the original contribution. An adequate contrastive linguistic description of the grammatical categories and coding of aspect in Czech and English that engages with relevant scholarship is lacking. There are some remarks about aspect in Chapter 4, but these are quite sparse and not sufficient for a reader who does not know Czech (or Slavic) aspect. There are some more remarks on page 195 of Chapter 6, but these are problematic (there are assumptions based on Russian that are oversimplified and not necessarily applicable to Czech). The first paragraph of the conclusion (p. 176) seems to suggest that "aspectuality" and "telicity" are different things, but telicity is usually understood as part of aspect. On p. 177 JJ suggests that it would be useful to try to "teas[e] apart the factors of aspect and directedness". However, directedness is part of aspect, so maybe it does not make sense to try to separate the two.

In addition to grammatical aspect, which can be operationalized because it is overtly marked, several other "eventuality" categories were included that perhaps have caused problems with the statistical models because a) they are collinear with aspect, and b) they are harder to operationalize. Collinearity is the bigger of the two concerns. For example, Vendlerian classes align with aspect: accomplishments and achievements tend to be perfective, while activities and states are imperfective, but the differences within each of these two pairs are very hard to operationalize for spontaneous language. "Directedness" also tends to align with perfective, which tends to be resultative and therefore directed.

III. Detailed evaluation of the dissertation and its individual aspects

1. Structure of the argument

The structure of the argument is mostly good, but sometimes there are inconsistencies or steps that were left out. For example, why exactly were only stroke/beat types of gesture included in the first study? And why do we get so much description of 63 silent videos of gestures were made for the second study when it later (p. 175) seems that only two of them were used for the test items (though presumably some of the others were used for distractors)? How were the sentences for the experiment constructed? Most of the sentences are rather vague, but the first one contains a long and unusual word *Uruguayci* 'Uruguayans' that might give readers a reason to pause. The statistical analyses present many visualizations but not enough discussion to provide a clear message for the reader. Is it the case that only one condition was tested for each pair of sentences? How was it decided which sentence would get assigned to which condition? And why then weren't the participants randomly assigned to two different groups so that each pair of sentences could be tested for both conditions? [Or maybe as a reader I just found it too hard to follow exactly what was in the experiment, in which case it could have been clearer?] For both studies, it would have been very helpful to get a summary of the data that showed the breakdown of how many observations there are for all of the values of the factors.

2. Formal aspects of the dissertation

The formatting is overall good, but there are a lot of small problems with the text (typos, errors with English articles, and the like). Most pages have several errors, although most of these do not hamper readability. There are also many sentences that are too long and/or have too many parentheses or are otherwise not very reader-friendly. I have marked up a paper copy with editorial comments on every page and will be happy to scan and send this in case JJ wants to publish all or part of this dissertation. In most of Chapter 5 there was a problem with the formatting of footnotes: many of the footnotes are labeled with the number 96 and have no reference in the text.

In many places the discussion is very abstract and few illustrative examples are provided. Here are some places where examples would have been helpful: Table 2 on p. 105, the stroke and predicate type combinations on p. 115, and all through 5.3 where it is hard to get a real sense of the data from visualizations alone. Sometimes terminology is inconsistent. The meetings that comprise the two corpora in the first study are initially referred to as "business meetings" on p. 110, but then on p. 111 we find out that these are meetings of faculty and students (not CEOs and businessmen), and in Table 3 on p. 113 they are termed "academic meetings".

I believe there is a misstatement on p. 159: "All items were simple transitive sentences with SOV or OSV word order and in the past or present tense." Of the 24 sentence pairs, 21 are past tense and 3 are future. None of them are present tense.

I believe there is a misstatement on p. 172: "The ended-ipfv responses were also characterised by the fastest RT." The ended-ipfv responses had the slowest RT, 3941ms; see Table 28 on p. 170.

The list of verbs and imageability ratings in Appendix E should have been arranged according to mean imageability instead of alphabetically.

Concerning the data "reduction" due to skewed reaction times in Chapter 6: Just because an observation is an outlier doesn't mean that we can discard it, there has to be a theoretical justification. Why not log-transform all the reaction times if they are so skewed? And: if the choice

disappeared from the screen after 10 seconds, how can it be that there were any responses after more than 10,000 milliseconds?

In the experimental study, why is there a comparison to a model without the one significant variable? Usually one tries to find an optimal model that retains only the significant variable. Including non-significant variables in a model reduces the model's efficacy.

3. Use of sources and/or material

Control of previous scholarship is particularly strong in this dissertation. However, JJ does not do enough to summarize and highlight the concepts most needed for the studies in the dissertation and does not return to relevant scholarship in discussing the results.

4. Personal contribution to the subject

There is a good statement of the contributions of this dissertation on page 6.

IV. Questions for the author

How and why are the results in this dissertation different from those found by Cienki & Iriskhanova (2018), who found that gesture did not align with aspect in German and Russian?

In the theoretical introduction, we read on pp. 47-48 a critique of the modularity of generative grammar, but on pp. 58-61 models by Levelt and others are proposed, and on p. 58 we read "The core idea behind Levelt's model is the hierarchical activation of separate modules". How can we reconcile these different approaches that alternatively eschew and promote a modular understanding of language and communication?

At the end of Chapter 4 (p. 108) JJ offers some "assumptions (not hypotheses)". All of these assumptions are basically the same: alignment of the expression of "eventuality" and gesture. But do we have to make this assumption? Could it be that gestures can instead complement grammatical aspect, highlighting features that are not coded in the speech stream?

In response to the comment on p. 129 that because a CART analysis was used instead of regression, it was not possible to include the speakers as random effects: Given that there weren't very many speakers in either corpus, why not just include the speakers as a factor in CART and then see whether/how CART separates them?

Concerning the discussion of various factors that potentially impact reaction times (mismatching conditions, sensitivity of participants, etc.): do we really know what the reaction times mean? Is there any straightforward way to interpret reaction time?

Concerning congruent vs. incongruent stimuli for the experiment in Chapter 6: Given that all possible combinations of aspect and gesture type are attested in the corpus (see Table 20 on p. 154), how can we really claim that some combinations are congruent vs. incongruent? Aren't all of them acceptable? Maybe to varying degrees?

V. Conclusion

This is a fascinating dissertation which I read with great interest. My many questions and critiques should be understood as a measure of how engaging I found this work to be.

I recommend the submitted dissertation with the tentative grade of pass.

April 2, 2021