

**The Prosodic and Information Structural Underpinnings of
Speech and Co-Speech Gesture Alignment in Hungarian**

by

Csilla Tatár

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in the University of Michigan
2026

Doctoral Committee:

Professor Ezra Keshet, Co-Chair
Professor Jelena Krivokapić, Co-Chair
Associate Professor Natasha Abner
Professor Dr. Frank Kügler, Goethe Universität Frankfurt
Professor Eric Swanson

Csilla Tatár

cstatar@umich.edu

ORCID iD: 0009-0006-6528-9505

© Csilla Tatár 2026

Dedication

To my parents and grandparents, whose hard work shouldered my education.

Acknowledgements

As the saying goes, *it takes a village...*, and it absolutely did.

Beyond those named below, there are inevitably others who contributed to my academic development, and I apologize for not being able to acknowledge everyone by name.

Central to my experience at UMich and to the completion of this dissertation are two remarkable scholars whom I am fortunate to call my mentors: my co-chairs, Ezra Keshet and Jelena Krivokapić. I am grateful, immeasurably so, for their generosity, support, encouragement, and patience at every step of the project. Ezra, thank you for taking me on as your advisee and for guiding me through these {great, challenging, easy, incredible, formative, ...} years. Your instruction, feedback, and calm approach to every concern kept me on course. Jelena, thank you for welcoming me into the world of prosody and for taking me on as a late advisee. I am grateful for the truly extraordinary care and attention you have invested in me and in this dissertation.

Due appreciation is owed to the members of my committee, Natasha Abner, Frank Kügler, and Eric Swanson, whose expertise and always constructive and encouraging feedback strengthened this project along the way. Thank you for your time, support, and thoughtful comments. I have benefited immensely from your questions and insights, and I am grateful for the opportunity to have learned from each of you.

Especially helpful to bringing the studies to life were Jungyun Seo, Yoonjeong Lee (Cynthia), Ruairidh Purse, and Yue Liang. I am indebted to all of you for your time, advice, and help during the data collection and processing stages. Thank you, Ruairidh, for training me on the

EMA and for helping me pilot, thank you Cynthia and Yue for your help with the data streams, your expertise was crucial to this project. And thank you Jungyun for *literally everything* – you did it all and more to help me from start to finish, and I could not be more grateful.

For her advice regarding collecting Hungarian data in an English-speaking environment, I am grateful to Savithry Namboodiripad. Thank you, Savi, for sharing your expertise with me and for giving me a fresh perspective.

Gratitude is owed as well to my participants (whom I will not name to preserve their anonymity). Should this dissertation reach you, please know that I am thankful for your time and the genuine effort you put into our data collection sessions. Without your commitment, this dissertation would not have been possible.

Huge thanks as well to Olawale Akingbade for being my EMA pilot participant! Thank you for your time and patience while we worked out the kinks.

I am grateful for all the help I received from the administrative staff of the Linguistics department, among them Jennifer Nguyen, Elexis Gillian, and Carl Abrego. Thank you for your assistance at many points throughout the years.

Joining Michigan meant joining a phenomenal academic community, and I will absolutely fail in naming everyone that has influenced me. Beyond those mentioned above (who were more directly involved with different aspects of the dissertation), I am especially thankful to Jonathan Brennan for his invaluable mentorship throughout the years. Thank you, Jon, for teaching me about conducting research, for guiding me through my first ever experiment, and for showing me what it means to be an educator. I have learned so, so much from you and I am truly grateful. Next, I am thankful to other amazing members of the linguistics faculty whose example

shaped my thinking and whom I greatly admire. Pam Beddor, Andries Coetzee, Jessi Grieser, Lisa Levinson, Steve Abney, and Acrisio Pires, thank you!

Key to my experience in graduate school have been the graduate students and alumni as well. Thank you, Jungyun Seo (*again*), Aliaksei Akimenka (alumnus), Lucy (Yu-Chuan) Chiang (alumna) for the heart-to-hearts and boba teas. Thank you to L. R. ‘Nik’ Nikolai, Sophie Faircloth, Olawale Akingbade (*again*), Kendall Lowe, Wyatt Barnes, Promise McEntire, Emma Portugal, Elliot Stork, Junyuan Zhao, Demet Kayabaşı, Isaias Ceballos III, Daniel Morales, Martin Mössmer, Nour Kayali, Sid Bhushan, Sovoya Davis, Cecilia Solís-Barroso, Anna Whitney, Yueke Zeng, Sylvan Jesień, and Rachel McCullough. Thank you to alumni Dominique Canning, Yourdanis Sedarous, Kelly Wright, Wilkinson Gonzales (my recruitment host), Sophie Eakins, Jeonghwa Cho, Laretta Cheng, Justin Craft, Tamarae Hildebrandt, Tzu-Yun Tung, Andrew McInnerney, and Yushi Sugimoto. I admire each of you for your dedication and brilliance.

Looking further back, I am thankful for the formative experiences at Eötvös Loránd University and elsewhere. First and foremost, I remain indebted to the late Professor Ádám Nádasdy, who first sparked my interest in linguistics during his unparalleled lectures and seminars and who propelled my learning in my BA and MA years at ELTE. Thank you as well to Ádám Endre Hamvas, Péter Mekis, Péter Szigetvári, Beáta Gyuris, Miklós Törkenczy, András Máté, Zsófia Zvolenszky, Judit Szalai, Krisztina Szécsényi, and Mark Newson for the many ways in which you supported me in my studies.

My deepest gratitude, finally, goes to my family. My parents have been my greatest champions and sources of support. Their love and unwavering belief in me mean everything, and I cannot thank them enough for all they do. Köszönöm, anya és apa. Next, thank you to Connor

Toole, my husband. Connor, I cannot begin to thank you for all your support (and all the coffee) from day one. Few (if any) are as lucky as I am to be in a *kayak* with someone like you. I would also like to thank my brother, Dávid, and his family, Ági and Bence, who always provided me with the best reasons to step away from work for a bit. Thank you as well to the Toole family, Denise, Mark, Cait, Brendan, and Emily (and baby James!) – thank you all for welcoming me into your home, thank you for the Thanksgiving dinners, and for looking out for me on your side of the pond.

This dissertation and the completion of my degree would not have been possible without all your support, thank you.

Table of Contents

Dedication	ii
Acknowledgements	iii
List of Tables	xii
List of Figures	xiii
Abstract	xvi
Chapter 1 Introduction	1
1.1 Introduction: multimodal communication	1
1.2 Co-speech gestures.....	3
1.3 Co-speech gestures and prosody	5
1.3.1 Temporal alignment	6
1.3.1.1 Alignment with lexical and phrasal prominence	6
1.3.1.2 Alignment within the Prosodic Hierarchy	8
1.3.2 Prosodic prominence and boundary effects on co-speech gestures	12
1.3.2.1 Prosodic prominence effects	12
1.3.2.2 Prosodic boundary effects.....	15
1.3.3 Theoretical approach within the study: Articulatory Phonology	17
1.4 Co-speech gestures and information structure	17
1.4.1 Information structure	19
1.4.1.1 Topic and Comment.....	20
1.4.1.2 Focus	29
1.4.1.3 Information status	46

1.4.2 Co-speech gestural marking of information structure	49
1.4.2.1 Focus and topic	50
1.4.2.2 Information status	54
1.5 Research Questions	57
1.5.1 Speech and co-speech gesture coordination	57
1.5.2 Prosodic boundary and prominence effects	63
1.5.3 Proposed study and dissertation chapters	66
Chapter 2 Methodology	67
2.1 Participants.....	67
2.2 Materials	68
2.2.1 Experimental conditions	68
2.2.1.1 Stimuli: Information structure.....	70
2.2.1.2 Stimuli: Prosodic boundary conditions	73
2.2.2 Tasks	74
2.2.3 Data collection procedure	76
2.2.4 Data processing and annotation	78
2.2.4.1 Constriction and pointing gestures.....	78
2.2.4.2 Prosody	81
Chapter 3 Coordination.....	83
3.1 Introduction and research questions.....	83
3.2 Methodology	84
3.2.1 Experimental conditions and stimuli	85
3.2.2 Data analysis	86
3.2.3 Statistical analysis.....	88
3.3 Results.....	90

3.3.1 Smallest variance of speech and co-speech gesture intervals [RQ1].....	90
3.3.2 Effects on coordination [RQ2].....	93
3.3.2.1 Effect of information structure.....	93
3.3.2.2 Effect of f0 contour type.....	94
3.3.2.3 Interaction of information structure and f0 contour type.....	96
3.4 Discussion.....	97
3.4.1 Coordination [RQ1].....	97
3.4.1.1 Coordinating landmarks in Hungarian.....	97
3.4.1.2 Contextualization.....	99
3.4.2 Effects on coordination in Hungarian [RQ2].....	102
3.4.2.1 Information structure.....	102
3.4.2.2 F0 contour and information structure.....	104
3.4.2.3 Contextualization.....	105
3.5 Conclusion.....	106
3.5.1 Broader implications.....	106
3.5.2 Limitations and future research avenues.....	106
Chapter 4 Boundary and Prominence Effects.....	109
4.1 Introduction.....	109
4.2 Methodology.....	112
4.2.1 Stimuli.....	112
4.2.2 Data analysis.....	113
4.2.2.1 Data labeling.....	113
4.2.2.2 Boundaries.....	114
4.2.2.3 Co-speech gesture position relative to target.....	116
4.2.2.4 Boundary and prominence related measures.....	119

4.2.3 Statistical analysis	121
4.3 Results	123
4.3.1 Initial-boundary effects [RQ3]	123
4.3.2 Information structure effects [RQ4]	128
4.3.3 Interaction model (boundary type and information structure) [RQ5]	134
4.3.3.1 Interaction of Boundary and Information structure: C1	135
4.3.3.2 Interaction of Boundary and Information structure: V1	136
4.3.3.3 Interaction of Boundary and Information structure: pointing	140
4.3.4 Post-hoc: Post-target boundary effect	140
4.3.5 Post-hoc: Boundary and Information Structure effects on post-target pause	141
4.4 Discussion	145
4.4.1 Prosodic boundary effects [RQ3]	145
4.4.2 Information structure effects [RQ4]	149
4.4.3 Interaction effects [RQ5]	152
4.5 Conclusion	154
Chapter 5 General Discussion	155
5.1 The results: an overview	158
5.1.1 Coordination	158
5.1.2 Temporal properties of pointing gestures	161
5.1.2.1 Prosodic boundaries	161
5.1.2.2 Prominence	163
5.2 The results: an Articulatory Phonology perspective	165
5.2.1 Coordination	165
5.2.2 Temporal properties	169
5.3 Language production and temporal alignment	172

5.4 Notes on Hungarian prosody	178
5.5 Limitations and future directions	182
5.6 Conclusion	186
Bibliography	187

List of Tables

Table 1 Experimental conditions. Dark gray shading indicates that no data was collected for the cell. Light gray shading indicates conditions not examined in the dissertation.....	69
Table 2 Stimuli conditions: green indicates conditions analyzed in the dissertation, yellow indicates data collected not as part of the dissertation, and gray shading indicates no data was collected for the cell.....	71
Table 3. Experimental blocks.....	75
Table 4. Pointing gesture dimensions labeled.....	79
Table 5. Pairs of landmarks considered for coordination. (Note that the numbers in the coding following the words ONSET and TARGET are for ordering purposes and are not meaningful.)	87
Table 6. Residual variances of the intervals.....	90
Table 7. Relative variance ratios relative to TARGET2__V1_FING_z (the baseline group) and results of bootstrapping (confidence intervals). Nonsignificant results are shaded light gray.	91
Table 8. By-participant raw median (ms) and variance (ms) of the intervals between the statistically not significantly different speech and co-speech gesture landmarks.....	93
Table 9. Number of tokens from information structure conditions in each f0 contour cluster. ...	96
Table 10. Measured lags for gauging gesture position relative to the target word	117
Table 11. Speech and co-speech gestural landmarks. (Note that the prefixes in the interval coding scheme (p1_, p2_, a,1_, a2_, etc.) are not meaningful.)	120
Table 12. Linear model outputs (significant and marginal effects of boundary on C1)	124
Table 13. Linear regression output of significant interactions of IS_SUBTYPE and BOUNDARY on C1, V1, and pointing movement durations.	134
Table 14. Pairwise comparisons of information structure effects on the post-target word pause duration (significant and marginal effects).	144

List of Figures

Figure 1 The Prosodic Hierarchy, adapted from Grice (2006).	9
Figure 2 Coupling relationships in a CVCV word between articulatory (C, V), tone (H/L), and co-speech (G) gestures. Solid lines indicate in-phase coupling, dotted lines indicate anti-phase coupling (see Goldstein et al., 2009; Nam, Goldstein, and Saltzman, 2009; Mücke et al., 2012). Red lines with a question mark indicate an unknown coupling relationship.	61
Figure 3 Potential effect of a preparatory tone on tonal alignment given a word boundary preceding the target (vs. no effect following a bigger boundary).	63
Figure 4 Schematic representation of the scope of the π -gesture's lengthening effect on co-active gestures. Darker to lighter shading indicates the decrease in effect. Figure adapted from Byrd & Krivokapić (2021).	64
Figure 5 Experimental setting	74
Figure 6. Experimental setup	77
Figure 7. Speech and co-speech gesture labeling in mvview: spectrogram, tongue dorsum (TD, vertical) lip aperture (LA), right index fingertip (FING, vertical). The target word ['mimɒ] is highlighted in green. The left end of the box (label) is the gesture onset; the left end of the shaded box is the gesture nucleus onset (onset of plateau); the dotted line is the gesture target; the right edge of the shaded box is the gesture nucleus offset; the right edge of the box is the gesture offset.	81
Figure 8. Pitch accent and f0 labeling convention. Panel (a) shows a token with a H pitch accent, where the PA and f0 labels coincide. Panel (b) shows a token with a L pitch accent, where the PA and f0 labels diverge.	82
Figure 9. Boxplots and scatterplots of each interval (standardized)	90
Figure 10. Variance ratios with bootstrap 95% confidence intervals relative to the interval with the smallest variance (TARGET2__V1_FING_z).	92
Figure 11. The most stable interval (TARGET2__V1_FING_z) across information structure conditions, ordered from smallest to greatest IQR.	94
Figure 12. Target word f0 contour clusters.	95

Figure 13. The most stable interval (TARGET2__V1_FING_z) across F0 contour clusters ordered from smallest to greatest IQR.	95
Figure 14. The most stable interval (TARGET2__V1_FING_z) across information structure conditions and f0 contour clusters.	97
Figure 15. π -gesture effect scope. Figure adapted from Byrd & Krivokapić (2021).....	110
Figure 16. Raw values of pause duration before the target word, separated by boundary type and participant.....	115
Figure 17. Raw values of pause duration after the target word, separated by participant.	116
Figure 18. Co-speech gesture position relative to the target word, separated by participant and boundary condition. The co-speech gestural timepoints were subtracted from the articulatory landmarks' timepoints (in ms). Negative range indicates that the co-speech gestural landmark follows the articulatory landmark.	118
Figure 19. Schematic representation of gesture landmarks.	119
Figure 20. Speech and pointing gestural regions of interest.	120
Figure 21. Initial-boundary effect on C1, separated by participant. Significance is indicated by asterisks, dots (.) indicate trends.	125
Figure 22. Initial-boundary effect on V1, separated by participant. Significance is indicated by asterisks, dots (.) indicate trends.	126
Figure 23. Initial-boundary effect on pointing phases, separated by participant. Significance is indicated by asterisks, dots (.) indicate trends.	128
Figure 24. Marginal effect of information structure on two V1 movements (P4).	129
Figure 25. Significant pairwise contrasts between model estimated marginal means of pre-target pause duration across information structure conditions. Significance is indicated by asterisks.....	131
Figure 26. Significant pairwise contrasts between model estimated marginal means of pointing onset to target duration across information structure conditions. Significance is indicated by asterisks.....	132
Figure 27. Significant pairwise contrasts between model estimated marginal means of pointing plateau and target to offset durations across information structure conditions. Significance is indicated by asterisks.	133
Figure 28. Boundary and IS_subtype interaction effect on C1 movement durations separated by participant.	137

Figure 29. Boundary and IS_subtype interaction effect on V1 movement durations separated by participant.	138
Figure 30. Boundary and IS_subtype interaction effect on pointing movement durations separated by participant.	139
Figure 31. Significant effects of post-target pause on pointing durations. Note, the effect on [P1] p3_point_re is only marginally significant (p=0.08).	141
Figure 32. Raw values of post-target pause durations separated by speaker and initial boundary condition.	142
Figure 33. Post-target pause duration separated by participant and pre-target boundary condition.	143
Figure 34. Post-target pause duration separated by participant and information structure condition.	144
Figure 35. Schematic representation of the co-speech gesture coupling with a tone gesture (A) or with a constriction gesture (B). Notation: C=consonant, V=vowel, T=tonal gesture, G=co-speech gesture.	166
Figure 36. Schematic representation of the co-speech gesture coupled with a tone gesture, which itself is coupled with a μ -gesture (A), and of both the tone gesture and the co-speech gesture coupled with a μ -gesture (B). Notation: C=consonant, V=vowel, T=tonal gesture, G=co-speech gesture, μ =temporal modulation gesture.....	167
Figure 37. Schematic representation of π -gesture effects in the IP and word boundary conditions.	171
Figure 38. Adapted from Kita & Özyürek (2003) to highlight areas of focus for prosody (in the Formulator) and information structure (in the Discourse Model).	174

Abstract

An extensive body of research has documented that speech accompanying gestures (co-speech gestures) are timed with speech in systematic ways, underscoring the multimodal nature of communication. Prosodically and pragmatically prominent units of speech have been shown to serve as anchoring points to co-speech gestures, and prosodic structure has been demonstrated to influence the temporal properties of co-speech gestures. Despite the accumulation of evidence, outstanding questions remain. The precise nature of alignment between the modalities and the roles of prosody and information structure are not yet understood, and the cross-linguistic examination of the effects of prosodic structure and prominence on gestures is limited.

In an effort to better understand the linguistic principles underlying speech and co-speech gesture coordination, the dissertation examines the respective roles of prosodic structure and information structure in speech and co-speech gesture alignment in Hungarian, a language in which information structure is expressed via word order modulation primarily. Two main sets of questions are explored: (1) What landmark of pointing gestures coordinates with what landmark of speech in this language, and is that coordination affected by information structure and intonation? (2) Do prosodic boundaries and prominence affect the temporal properties of pointing gestures, as they do speech gestures?

These research questions are addressed in two semantically-pragmatically controlled experiments during which kinematic, visual, and acoustic data were collected. The results of the first study establish that the word-initial lexically stressed syllable (specifically, the vowel)

anchors pointing gestures in Hungarian. The stability of this coordination is not significantly affected by information structure, intonation contour, or their interaction. The second study presents some evidence that prosodic boundaries can affect the temporal properties of gestures. Additionally, although prominence has no temporal effect on consonants and vowels, an effect is observed on pointing gestures, likely due to information structure modulating boundary magnitude, and thus pointing durations. Altogether, serving also as a mediator between information structure and gesturing, prosodic structure appears to drive the synchrony between speech and co-speech gesture realization in Hungarian. The results are interpreted within the framework of Articulatory Phonology.

The conclusions broadly corroborate prior research on gesture behavior in other languages, in that the co-speech gesture coordinates with the prominence-bearing syllable, although language-particular patterns are also observed, which likely relate to the prosodic typology of the languages so far studied.

Chapter 1 Introduction

1.1 Introduction: multimodal communication

In interpersonal communication, multiple channels are employed to construct and convey messages. One channel includes semantic information pertaining to the lexical content of the message within its given context. Another channel pertains to prosodic information, which refers to characteristics such as prominence relations and phrasing in an utterance. A third channel involves gestural information, referred to as co-speech gestures, that complement and/or enhance the message, including hand gestures and eyebrow movements. These channels converge seamlessly in communication, and the study of their coordination has emerged as a significant and promising area of research, following Kendon (1972, 2004) and McNeill (1992, 2005). How these different dimensions co-influence the temporal alignment and properties of co-speech gestures has been the subject of more recent investigations. The present project picks up this thread of research by examining how information structure, prosodic structure, and co-speech gestures interact in Hungarian.

Landmarks of speech prosody and co-speech gestures are known to systematically co-occur or *align* (see e.g. Gregori & Kügler, 2026; Kügler & Gregori, 2023; Rohrer, 2022, Shattuck-Hufnagel & Ren, 2018; Fung & Mok, 2018; Krivokapić et al., 2017; Esteve-Gibert & Prieto, 2013; Loehr, 2012; Mendoza-Denton & Jannedy, 2011; Leonard & Cummins, 2011; Swerts & Krahmer, 2010). The studies report on (variably defined) temporal alignment between co-speech gestures and pitch accents. However, coordination studies often elicit prosodic

prominence with focus-stimuli (expecting a pitch accent to occur on the focused constituent in the object languages) or do not specify the type of (pragmatic) prominence examined, potentially conflating their effects.

Pragmatic prominence does, however, have an effect on alignment. Türk (2020) observes that co-speech gestures co-occur with *focus* and *topic* more than with *background* and concludes that *background* elements do not have discursive prominence, which might entail that co-speech gestures are not attracted to them. In contrast, Kügler & Gregori (2023) do observe a relatively large proportion of co-speech gestures to co-occur with *background* elements. Importantly though, the authors report pitch accents under focus to more closely align with co-speech gestures than non-focused (background) pitch accents. It is unclear, therefore, to what extent the systematic alignment observed in many studies is due to prosodic prominence (mainly pitch accentuation), focus, or specifically focus-related prominence.

This introductory chapter is organized as follows: **Section 1.2** gives a brief overview of co-speech gestures; **Section 1.3** reviews the relevant literature on the role of prosody in speech and co-speech gesture alignment (reviewing prominence and boundary related effects on temporal alignment and co-speech gesture realization); **Section 1.4** gives an introduction to concepts of information structure, with a special emphasis on their structural and prosodic expression in Hungarian, and reviews the literature on the co-speech gestural marking of information structure. **Section 1.5** combines insights from the previous sections and details the research questions addressed in the dissertation.

1.2 Co-speech gestures

The present subsection serves to introduce co-speech gestures briefly; for a thorough overview of the topic, see Abner et al. (2015). Manual co-speech gestures are hand movements that accompany speech; Kendon (2004) defines them as visible actions “used as an utterance or as a part of an utterance” (2004:7). Co-speech gestures are meaningful, but they vary in their relation to the semantic meaning of the utterance they accompany (McNeill & Duncan, 2000). A co-speech gesture might add no new information to an utterance, or it might complement or supplement the utterance, and it might replace part of the utterance entirely (Bergmann et al., 2011). In McNeill and Duncan’s (2000) terminology, the co-speech gesture and its *lexical affiliate* (the expression co-occurring with a gesture) are “co-expressive”, in that they express the same underlying concept. A model proposed in Kita & Özyürek (2003) captures co-speech gesture production in relation to aspects of speech production by stipulating that conceptual and linguistic information relevant to speech production can be available to co-speech gesture production. The temporal alignment of co-speech gestures relative to their lexical affiliate is an empirical question (and is not addressed in the Kita & Özyürek (2003) model), but conceptually they correspond to the same part of the utterance. This property of co-speech gestures is important to note here, as the dissertation examines co-speech gestural alignment in part as a function of the role of the speech unit a gesture co-occurs with in Hungarian.

McNeill (1992, 2005) categorizes co-speech gestures as *iconic*, *metaphoric*, *deictic*, and *beat-like*. Iconic gestures represent concrete objects, metaphoric ones represent abstract ideas, deixis refers to “spatial relations with concrete or abstract entities” (such as pointing), and beats have discourse-pragmatic highlighting functions in accordance with speech prominence and rhythm (Rohrer, 2022). These categories are not discrete, however; rather, they may be thought

of as dimensions, with co-speech gestures having both referential (meaning-related) and rhythmic (beat-like) properties (McNeill, 2005, 2006; Shattuck-Hufnagel & Prieto, 2019; Rohrer, 2022). The dissertation focuses on deictic (pointing) gestures.

Co-speech gestures are composed of *gestural phases*, which may be thought of as parts of the articulatory process of gestures. The following descriptions of the phases are summarized from Wagner et al. (2014). *Rest position* is the stable position from which a gesture is started. *Preparation* is the movement between the rest position and the place from where the meaningful part of a gesture begins (Kendon, 1972). Between the preparation phase and the meaningful part is a potential *hold* (called *pre-stroke hold* in Kita, 1990), when movement is suspended. The meaningful part is called the *stroke* of a gesture (Kendon, 1972); this phase is produced with the most effort and it is the only obligatory part of a gesture. At the end point of the stroke is the so-called *apex* (defined as maximum displacement and zero velocity). If the gesture is held in place past the apex, the phase is called a *post-stroke hold* (Kita, 1990). Finally, the gesture concludes with *retraction*, or return to rest position (Kendon, 1972). A gestural phase cycle is called a *gestural phrase* (Kendon, 1972). Gestural phrases are organized into *gestural units*, which are demarcated by two rest positions (for further elaboration, see Kendon, 1980 and Türk, 2020). That is, a gestural unit can be composed of one or more gestural phrases, and a gestural phrase is composed of the gestural phases (for instance, a pointing gesture consisting of the phases of a stroke, hold, and return can be repeated multiple times, with each stroke-hold-return sequence constituting a phrase). In examining speech and co-speech gesture alignment, the dissertation also asks which part(s) of a gesture most closely and most stably align with which part(s) of a speech unit in Hungarian, and in examining the temporal properties of gestures, the dissertation asks which phases of gestures may be influenced by linguistic structures.

McNeill (1992) posits the existence of three synchrony rules along which speech and gesture coordinate: the *phonological*, *semantic*, and *pragmatic* synchrony rules. The next subsection reviews the literature on the temporal alignment of co-speech gestural and prosodic landmarks, providing – in essence – support for the phonological synchrony rule.

1.3 Co-speech gestures and prosody

Co-speech gestures and certain speech events have been observed to frequently occur together. Kendon (1972) is credited for the insight that co-speech gestures synchronize with stressed syllables, and McNeill (1992) introduces the *phonological synchrony rule*, which posits that the stroke of a co-speech gesture “precedes or ends at, but does not follow, the phonological peak syllable of speech” (McNeill 1992:26). Although the phrasing of the phonological rule is quite ambiguous, it has been tested many times, extending alignment with stressed syllables to alignment with phrasal prominence, as well as alignment between prosodic and co-speech gestural phrases of various levels.

The present section reviews the relevant literature that shows compelling evidence for the tight connection between prosody and co-speech gesture, reviewing in *subsection 1.3.1* the temporal alignment patterns and in *subsection 1.3.2* the effect of prosodic prominence and boundaries on co-speech gesture realization.

1.3.1 Temporal alignment

1.3.1.1 Alignment with lexical and phrasal prominence

1.3.1.1.1 Lexical stress

The relationship between *lexical stress* (i.e. the relative emphasis of a syllable in a word, e.g. *REcord*, the noun and *reCORD*, the verb) and co-speech gesture coordination was investigated in Rochet-Capellan et al. (2008). Although the authors report differences in the exact alignment, they find tight temporal coordination between jaw movement in stressed syllables and the target of the pointing gesture overall. Fuchs & Reichel (2016) also examine the alignment of pointing gestures in counting out rhymes and report the turning points between two strokes to align with vowels frequently, and with consonants when the onset of the syllable is complex. Importantly, lexical stress can anchor co-speech gesture targets.

1.3.1.1.2 Phrasal prominence

Co-speech gestures have been shown to align with prosodically prominent expressions in languages in which phrasal prominence is expressed via pitch accentuation. *Phrasal prominence* refers to the relative emphasis on a word compared to its surroundings in a sentence (for a recent review, see Ladd & Arvaniti, 2023), *pitch accentuation* refers to the meaningful modulation of fundamental frequency (f_0); and *fundamental frequency* refers to the rate of vocal fold vibration, the perceptual correlate of which is called *pitch*. Pitch accents can be high (H), low (L), or their combinations, and have *tonal targets*, which are locally the highest (for H) or lowest (for L) points in f_0 . See Ladd (2008) for more information on pitch accents.

Studies find that co-speech gestures and pitch accents co-occur systematically (Gregori & K gler, 2026; K gler & Gregori, 2023; T rk & Calhoun, 2023; Rohrer et al., 2023; Shattuck-

Hufnagel & Ren, 2018; Esteve-Gibert & Prieto, 2013; Loehr, 2012; Leonard & Cummins, 2011; Mendoza-Denton & Jannedy, 2011; Roustan & Dohen, 2010; Swerts & Krahmer, 2010; Yasinnik et al., 2004). Often, the co-speech gesture apex (the gestural target) is reported to align with the pitch accented syllable (Gregori & Kügler, 2026; Kügler & Gregori, 2023; Türk & Calhoun, 2023; Leonard & Cummins, 2011; Roustan & Dohen, 2010; Loehr, 2012; Mendoza-Denton & Jannedy, 2011; Yasinnik et al., 2004). In the absence of a pitch accent, Türk and Calhoun (2023) observe an *accentless low tone* (L) on the left edge of *prosodic words* (which is a level of organization in prosodic structuring, see the next subsection, but also see Odden, 1986; Selkirk, 1978) to coordinate with co-speech gestures in Turkish. The underlying principles and the exact characteristics of speech and co-speech gesture co-occurrence, however, are not yet understood.

Note that co-occurrence has been variously defined in the literature, ranging from (a) any temporal overlap between the prominent words or phrases and co-speech gestures, (e.g. Carignan et al., 2024; Shattuck-Hufnagel & Ren, 2018), (b) co-occurrence within a specified amount of time, for instance, within the average duration of a syllable (e.g. Türk & Calhoun, 2024; Loehr, 2004), and (c) the amount and stability of the lag between co-speech gestural and prosodic landmarks, i.e. the smallest variance in distance between specific points of a co-speech gesture and a prosodic event (e.g. Leonard & Cummins, 2011; Katsika et al., 2014; Krivokapić et al., 2017). Although due to this variability in defining alignment, the findings in different studies cannot always be compared to one another in a straightforward manner, it is a well-supported observation that co-speech gesture targets are often aligned with prosodically prominent units.

If co-speech gestures indeed align with prosodically prominent units, as evidenced in the studies cited above, it is plausible that the manner in which they are aligned is mediated through

the prosodic system of the particular language. That is, in a language with pitch accents, the co-speech gestural target aligns with the pitch accent target, and in instances without a pitch accent, the co-speech gesture aligns with a different prosodic target, such as a prosodic edge. A confounding factor in alignment studies is that prosodic prominence is often elicited via focus. One question arising from this practice is whether the close alignment of the prosodically and pragmatically prominent speech units and co-speech gestures is due to focus in particular, increased prosodic prominence (regardless of the focused vs. not focused status of the lexical affiliate), or specifically focus-related prosodic prominence. That is, does the close alignment observed occur due to semantic-pragmatic factors, prosodic factors, or a complex interplay of the two? A further question emerging from research cited above is how speech and co-speech gestures align in a language, such as Hungarian, where the primary expression of pragmatic prominence is word order modulation, and where prominence may additionally be expressed via pitch accentuation (Genzel et al., 2015; Mády, 2015) and/or phrasing (i.e. the insertion of a pause before an expression to signal its prominence, see e.g. Mády et al., 2016; Mády & Kleber, 2010). The expression of pragmatically prominent units (focus and topic) in Hungarian will be detailed in *Section 1.4*.

1.3.1.2 Alignment within the Prosodic Hierarchy

Prosodic Hierarchy (Pierrehumbert & Beckman, 1986; Pierrehumbert, 1980) refers to the organization of speech into prosodic units. **Figure 1** below serves as an illustration, where the units range from syllable to utterance, and the smaller units are the building blocks of larger ones (for an overview, see Shattuck-Hufnagel & Turk, 1996 and Arvaniti & Fletcher, 2020). In **Figure 1**, the “Smaller Phrase” is a generic level between the Intonational Phrase and Word levels (Grice, 2006). The number of levels in the hierarchy is language specific. In Hungarian,

the *Accentual Phrase* (AP), a sequence between two pitch accents, has been suggested as a level between the Intonational Phrase and (Prosodic) Word levels (Benus, Reichel, & Mády, 2014; Mády et al., 2013). In English, an *intermediate phrase* (ip) has been suggested as a level between IPs and prosodic words (Beckman & Pierrehumbert, 1986); as pointed out in Mády et al. (2023), the presence of an ip level has not been established in Hungarian, and Varga (2010) argues against its presence in the language. For the purposes of the dissertation the important point to recognize is simply that larger and smaller prosodic units exist.

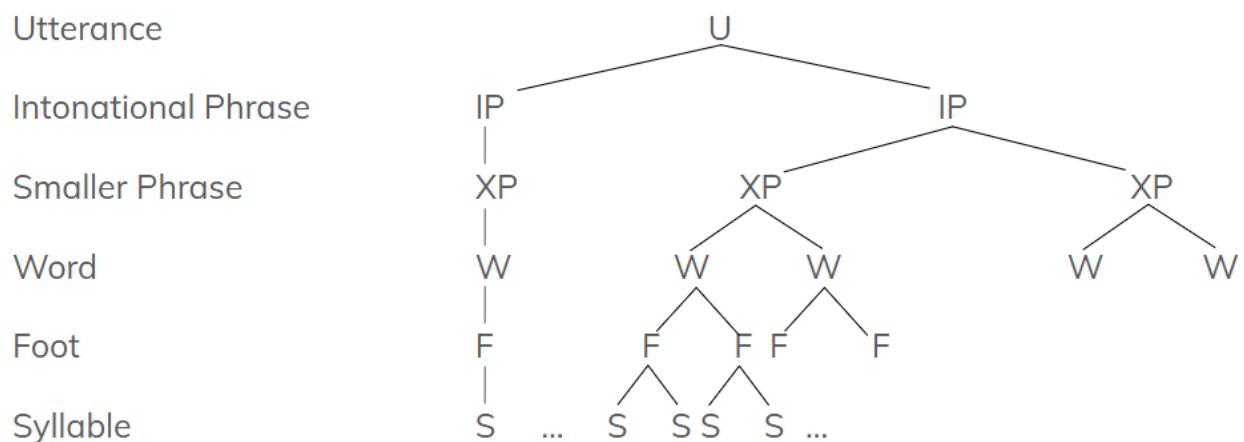


Figure 1 The Prosodic Hierarchy, adapted from Grice (2006).

Beyond prosodically prominent phrases, studies also report on higher level prosodic units to have an organizing effect on co-speech gestures as well (see Kendon, 2004). For instance, Loehr (2012) finds that co-speech gesture phrases show a tendency to align with intermediate phrases (ips), but not with larger intonation phrases (IPs) in English. Similarly, Shattuck-Hufnagel & Ren (2018) find non-referential rhythmic gestures to group into units which align with “higher-level prosodic constituents” in English academic-style speech (2018:10). Investigating alignment in French, Rohrer et al. (2019) report that non-referential rhythmic co-speech gestures (*beats*) are more likely to co-occur with phrase-final and phrase-initial pitch

accents than with medial ones. The authors suggest that the co-speech gestures may mark accentual phrases (APs). Rohrer et al. (2023) report that although around half of the observed co-speech gestures in French fall outside of the accented syllable, they occur *around* pitch accented syllables (84%), with many co-occurring with ip-initial pitch accents in particular (67%), which the authors interpret as phrase-initial strengthening. These sets of results show that co-speech gestures often occur at prosodically relevant parts in utterances, indicating an organizing role of prosodic structure in co-speech gesture production.

In examining alignment in French with a different criterion, Ferré (2010) finds that gestural strokes and phrases tend to begin before and end after the corresponding words and IPs, respectively. We may note that a tendency to start and end the co-speech gesture before and after the corresponding speech unit is not necessarily a case of *misalignment*; rather, it may be a specific type of alignment, referred to as *containment* in Türk & Calhoun (2024) who find a similar pattern in Turkish. The authors observe the occurrence of multiple intermediate phrases within one gestural phrase, where the left boundary of a starting ip in a group of ips aligns with the left boundary of a gesture phrase, and the right boundary of the last ip in the grouping aligns with the right boundary of the same gesture phrase (=containment). As described above, in examining co-occurrence patterns in Turkish, Türk & Calhoun (2023) find most co-speech gesture apexes to co-occur with pitch accents, and in the absence of pitch accents, to occur with an accentless low tone that marks the left edge of prosodic words. The authors interpret the results as an indication of co-speech gesture apexes being sensitive to prosodic phrasing, synchronizing with prosodic words. Further, the authors report an effect of prenuclear versus nuclear ip type found in the tightness of alignment between tonal events and co-speech gestural apexes, with a closer synchrony in nuclear ips. This result the authors interpret as

synchronization being sensitive to phrasal prominence as well. These results further indicate an organizing role of prosodic structure in the temporal alignment of speech and co-speech gestures, underscoring the relevance of prosody to co-speech gesture realization.

Prosodic boundaries between the prosodic units in the prosodic hierarchy have been suggested to affect the alignment of speech and co-speech gestures as well. Esteve-Gibert & Prieto (2013) show that in Catalan, both the pitch accent target and the pointing gesture apex occur sooner in phrase-final position (before a boundary), than in non-phrase final position. The authors interpret the results as stemming from the avoidance of tonal crowding: the pitch accent target occurs earlier phrase-finally because it precedes a boundary tone (a tonal event at prosodic boundaries), and the co-speech gestural peak occurs earlier because it is synchronized with the pitch accent. Again in Catalan, Esteve-Gibert et al. (2017) assess the effect phrase boundaries have on the alignment of head nods with focused prosodic words. This second study finds that head movements are co-extensive with their lexical target, with the left edge of the focused prosodic word synchronizing with the onset of the gesture stroke and the gesture peak showing a similar sensitivity to the following prosodic boundary as found in Esteve-Gibert & Prieto (2013).

The present subsection presented evidence for the close alignment of prosodic prominence (lexical and post-lexical) and co-speech gestural targets, as well as for the broad synchronization of co-speech gestural and prosodic phrases of varying levels. Studies thus highlight the fundamental role prosodic structure, both prominence and phrasing, plays in co-speech gesture production. The next section reviews the effect of prominence on co-speech gesture realization, as well as the handful of studies examining the effect of prosodic boundaries

(the edges of prosodic phrases of various levels in the prosodic hierarchy) on co-speech gestures and their alignment.

1.3.2 Prosodic prominence and boundary effects on co-speech gestures

1.3.2.1 Prosodic prominence effects

Pitch accented words have certain phonetic correlates: accented syllables tend to be longer, less coarticulated, and produced with greater intensity (for an overview, see van Heuven & Turk, 2020). Recent research has targeted the question as to whether these effects can also be seen on co-speech gestures that co-occur with pitch accents and indeed this appears to be the case.

Prominence effects are reported in Parrell et al. (2014), Rusiewicz et al. (2014), and Krivokapić et al. (2017). In an Electromagnetic Articulography (EMA) study, Parrell et al. (2014) find a close connection between the duration and magnitude of speech gestures and finger tapping gestures under cross-modal phrase-level prominence. The study asked participants to produce a single syllable (“ma” or “mop”) while tapping simultaneously. Participants were instructed to produce either the syllable or the tap with more emphasis, simulating prominence. The authors looked at the cross-modal effect of the emphasis; i.e., whether a stronger tap results in a more prominent syllable, and vice versa. Results show an effect of phrase-level prominence in both directions: the movement duration and magnitude are affected in both speech and co-speech gesture articulation when either the syllable or the tap is produced with more emphasis. Additionally, the authors interpret the regularity of the temporal lag between speech and gesture in both stressed and unstressed conditions as a tight coupling relationship between speech and gesture, which is further highlighted under prominence with a shorter lag. Thus, the co-speech

gesture is more prominent and more tightly coordinated with speech under phrase-level prominence. One note of caution is that the experiment did not examine the effect in sentences; therefore, although indicating a physical coupling relationship between speech and gesture, the results are not straightforwardly transferable to speech scenarios in which specific, linguistic prominence is at work.

Rusiewicz et al. (2014) report on the relative effect of contrastive pitch accentuation on pointing gesture duration in English. Prosodic prominence is elicited via question-answer pairs, in which the answer is either agreement or disagreement (the questions target a specific focus type to be produced in the answer, agreement or correction, as in this language, pitch accentuation is expected to mark focus). The study finds that the co-speech gesture is significantly longer when co-produced with a contrastive pitch accent (elicited via *corrective* focus).

In another EMA study, Krivokapić et al. (2017) examine co-speech gestures in controlled, but real sentences of English. Similarly to Rusiewicz et al (2014), prosodic prominence is elicited via question-answer pairs, in which the question targets a specific focus type to be produced in the answer. The types of focus targeted are *contrastive* (elicited via *correction*), *narrow*, and *broad* focus, as well as a deaccented baseline. The rationale behind using these types of focus is the observation that they differ in prominence strength in English (*contrastive: corrective* is expected to be more prominent than *narrow*, and *narrow* to be more prominent than *broad* focus). The authors report the pointing gesture to lengthen under prominence in a similar pattern to the lengthening of speech gestures, suggesting a tight relationship between the two.

Studies thus indeed support the idea of a tight coupling relationship between speech and gesture. In Krivokapić et al. (2017), the direction of prominence effect examined is from speech to co-speech gesture, while in Parrell et al. (2014), both directions (speech to gesture, gesture to speech) are studied, and the prominence effect appears to be bidirectional, which may raise the question as to the origin of the synchrony. Given the assumption of semantic and pragmatic synchrony, it is not counterintuitive that a co-speech gesture and its lexical affiliate are similar in their relative prominence levels too. Indeed, in a recent study on perceived prosodic and co-speech gestural prominence in different focus conditions in German and Catalan, Gregori, Sánchez-Ramón, Prieto, & Kügler (2024) find that the perceived prosodic prominence of focus types (*informational* < *contrastive* < *corrective*) aligns with the perceived prominence of co-occurring gestures.

One possibility explaining the cross-modal prominence effects is rooted in the biomechanics of gesturing. Studies find a strong coupling relationship between co-speech gestures and speech via the respiratory system (Werner et al., 2024; Kadavá et al., 2023; Pouw et al., 2020a; Pouw et al., 2020b; Pouw et al., 2020c). For instance, Pouw et al. (2020a) report that speech intensity is higher when co-occurring with arm or wrist movement as opposed to lower intensity when speech and movement alternate. The authors also find changes in f0 to correlate with arm movement due to changes in chest-circumference during movement. Relatedly, Pouw et al. (2020b) find greater mass movement (arm) to have a larger effect on intensity and f0 than smaller mass movement (wrist) in fluent speech. With evidence of both linguistic and biomechanical effects reported, it is unclear to what extent prosodic prominence (separately from biomechanics) affects co-speech gestures; that is, how much of the observed effects are due to specific linguistic factors.

1.3.2.2 Prosodic boundary effects

In addition to prominence, the question of prosodic boundaries affecting gestural realization has also been raised. *Prosodic boundaries* are the edges of prosodic phrases of various levels in the prosodic hierarchy. Boundaries between phrasal units are marked by certain temporal properties: articulatory gestures are longer, larger, and less overlapped at phrasal boundaries. This phenomenon is called *boundary-related lengthening*, and it has been shown in numerous studies for many languages, with the amount of lengthening increasing with *boundary strength* (for reviews, see Paschen et al., 2022; Byrd & Krivokapic, 2021; Krivokapić, 2014; Fletcher, 2010). As mentioned above, the number of levels in the hierarchy is not fixed, and boundary strength is to be understood relative to the levels of prosodic phrases in a given prosodic system. For instance, given a boundary related to an intermediate phrase (ip) and one related to an Intonation Phrase (IP), the IP boundary is said to be stronger and its effect greater, producing a greater amount of lengthening in the boundary-adjacent segments relative to lengthening related to the ip-boundary. The effect is strongest closest to the prosodic boundary and decreases further away from it. Relatively stronger boundaries can also be marked by pauses and boundary tones.

A theoretical framework with explicit predictions regarding the temporal effect of prosodic boundaries is the π -gesture model developed within the framework of Articulatory Phonology (Browman & Goldstein, 1992, 1995; Goldstein, Byrd, & Saltzman, 2006). In the AP framework, the units of both linguistic representation and physical action are *gestures* (Browman and Goldstein, 1989), and speech production is considered as the coordinated movement of the moving parts of the vocal tract, called *articulators*. Importantly, gestures can temporally overlap with one another, and the precise coordination of co-active gestures has been the subject of

research in Articulatory Phonology in pursuit of explaining various speech phenomena. Two stable coordination patterns suggested in AP studies are *in-phase* and *anti-phase*. In the former, two gestural events are observed to occur together (they begin simultaneously), while in the latter, they are observed to begin sequentially.

Research in the Articulatory Phonology framework has been extended to incorporate prosody with the proposal of prosodic gestures (for an overview, see Byrd & Krivokapić (2021) and Krivokapić (2014)). Three prosodic gestures have been proposed within the AP framework: π -gesture (Byrd & Saltzman, 2003), μ -gesture (Saltzman et al., 2008), and *tone gesture* (Gao, 2008). The π -gesture is a prosodic boundary gesture that has been proposed as an account for the boundary-related lengthening phenomena. The π -gesture thus does not have an *articulatory* realization of its own; rather, it has a temporal effect: the slowing down of gestures co-active with it. π -gestures can differ in magnitude with a proportionate difference in their effects (greater magnitude, greater effect). The model predicts co-active gestures to be affected by the π -gesture as a function of its activation strength and the amount of overlap between them (see Byrd & Saltzman, 2003). Next, a modulation gesture, called the μ -gesture, gives a model of the effects of lexical stress on constriction gestures (consonant and vowel gestures) Finally, the *tone gesture* is meant to capture the properties of tonal events and their coordination with other types of gestures.

An advantage of viewing and modeling speech as gestural coordination is the natural extension of research within the Articulatory Phonology framework to non-speech modalities of communication, such as co-speech gestures. The effect of prosodic boundaries in this spirit has been examined in Krivokapić et al. (2017), who demonstrate that the presence of prosodic boundaries indeed influences the relative temporal properties of co-speech gestures in English.

Using EMA data, Krivokapić et al. (2017) report phrase-final, but not phrase-initial lengthening to occur in co-speech gestures (pointing gestures, specifically) at prosodic boundaries. The authors attribute the lack of *initial* lengthening to the pointing gesture falling outside the effect of prosodic boundaries, an explanation rooted in the π -gesture model. Importantly, though, the study shows that the prosodic boundary effect on co-speech gestures parallels the effect on speech gestures. The results of this study thus indicate that co-speech gestures behave similarly to speech articulatory gestures under prosodic boundaries; that is, they lengthen, which suggests a close relationship between co-speech gestures and prosodic structure.

1.3.3 Theoretical approach within the study: Articulatory Phonology

The theoretical approach adopted in the dissertation is that of Articulatory Phonology, as introduced in the previous subsection. Adopting this framework involves a commitment to the view of speech and co-speech gesture co-production as gestural coordination (as discussed above), which bears on the methodological approach to and interpretation of the alignment question addressed in **Chapter 3**, as well as on the methodology and interpretation of prominence and boundary related temporal phenomena addressed in **Chapter 4**.

Following the review of information structure and co-speech gestures in **Subsection 1.4**, **Subsection 1.5** details the specific research questions and the related predictions as they arise from the Articulatory Phonology framework and the π -gesture model.

1.4 Co-speech gestures and information structure

Studies have reported on the role of *information structure*, including *information status* in the complex relationship between speech and co-speech gestures. Information structure pertains

to the organization of information within an utterance, such as highlighting the matter under discussion (referred to as *topic*), as well as identifying the most important or contrastive piece of information in an utterance (referred to as *focus*). *Focus*, and to a lesser extent *topic*, have been found to be frequent anchors of co-speech gestures (Türk & Calhoun, 2023, 2024; Kügler & Gregori, 2023; Lehečková et al. 2022; Türk, 2020, Ambrazaitis & House, 2016, 2017; Ebert et al., 2011, among others). Next, *information status* relates to whether (some of) the information relayed in an utterance has already been shared between interlocutors or is being introduced for the first time; that is, whether the utterance in question provides *new* or *given* information within the context of the conversation. Studies find that *new* information is more likely to be accompanied by a co-speech gesture than is *given* information (Rohrer, 2022; Debreslioska et al., 2013). These findings demonstrate the strong connection between information structure and co-speech gestures.

The present section surveys the theoretical and experimental literature on information structure (focusing on Hungarian), as well as the main threads of research on the relevance of information structure (IS) to speech and co-speech gesture alignment. The subsection is meant to introduce the expression of IS in Hungarian, as well as highlight the differences between Hungarian and languages studied thus far for co-speech gesture alignment in order to motivate the main research questions of the study; i.e., how the language specific expression of information structure may lead to differences in speech and co-speech gesture alignment. The section is organized in the following manner. **Subsection 1.4.1** introduces information structural notions with a focus on their semantic and pragmatic properties in Hungarian, along with notes on their expression in Hungarian in word order and prosody. **Subsection 1.4.2** reviews the

literature on the co-speech gestural marking of information structure and specifies how a study on Hungarian can continue this line of research.

1.4.1 Information structure

The present section gives an overview of information structure and its central notions, with notes on their structural and prosodic expression in Hungarian.

Following Chafe (1976), according to Krifka & Musan (2012), *information structure* (=IS) represents the way in which speakers package the message they wish to convey to their interlocutor, with whom they share a common ground. *Common ground* represents the insight that (groups of) interlocutors are aware of certain ideas and entities at a given time, or, more formally, common ground consists of a set of propositions that they all accept and a set of entities (*discourse referents*, in the sense of Heim, (2002 [1983])) that they have all been introduced to (Krifka & Musan, 2012; common ground as a *context set*: Stalnaker, 1978). In the question-answer pair given below in (1), (A) requests information, and (B) updates the common ground with the relevant information.

[1] Updating the common ground

A: *What is a good place to take some time off?*

B: *A Faraday cage (is a good place to take some time off).*

Common ground is not static; rather, as interlocutors keep exchanging information, the common ground keeps being updated. As Krifka & Musan (2012) point out, when constructing a message, speakers take into account not only what is already part of the common ground, but also what may be accommodated for the purposes of the conversation via accepting information

not explicitly stated as true. Central notions of IS are *topic-comment*, *focus-background*, and *givenness*.

Hungarian has been described as a *discourse configurational* language, in which discursive needs determine or at least strongly influence word order (É. Kiss, 1998; Varga, 2017; Surányi, 2016; É. Kiss, 2016). More precisely, Surányi (2016) defines discourse configurationality as being characteristic of “(...) languages in which a particular phrase structure configuration is systematically and exclusively associated with some Information Structural category falling under the notions of Topic or Focus (...)” (Surányi 2016: 422). In Hungarian, topic and focus in particular are said to occupy the specifier positions of functional projections in the left periphery of the sentence (É. Kiss, 2008). Word order, in broad strokes, is generally *topic + focus + verb + post-verbal region*.

1.4.1.1 Topic and Comment

To explain what *topic* and *comment* are, Reinhart (1981) proposes a metaphor of a library catalog (specifically, a subject catalog). In this proposal, propositions are classified into subsets stored in the common ground under a “defining entry”, or *topic*, and the set of propositions under an entry represents the interlocutors’ knowledge of that particular entry, thus storing the content of the *comment* part of a sentence (Reinhart, 1981:80). Krifka (2008) – following this metaphor – defines the topic constituent as identifying the entity or set of entities in the common ground of which new information is being shared and so entered into to the common ground, where it is stored as corresponding to the identified topic entity. Reinhart (1981) further specifies that topics are *referential* entries; such as NPs (or DPs). See (2) below, in which “Diogenes of Sinope” is introduced as the topic in A’s question, taken up again in B’s answer, which also provides the

requested information, which is then entered into the common ground under the entry “Diogenes of Sinope”.

[2] Topic and Comment

A: *What was Diogenes of Sinope looking for while holding a lantern in broad daylight?*

B: [TOPIC *He*] [COMMENT *was looking for an honest man*].

1.4.1.1.1 The syntactic topic position

Turning to Hungarian, in her discussion, É. Kiss (2008) argues that “(t)he topic is an XP extracted from the functionally extended verb phrase into the left periphery of the sentence” (pp. 288), proposing that it occupies the specifier position of TopP, a functional projection, which may be iterated given multiple topics. In terms of word order then, the Hungarian topic is assumed to appear before the comment (see (3) below).

[3] Schematic *Topic + Comment* word order:

Topic + [COMMENT Verbal modifier (VPX) + Verb + post-verbal elements]

A: “How are the buds in the garden?”

B: [TOPIC *A rügyeket*] [COMMENT *meg-csípte a dér (a kertben)*].

the buds.ACC VPX- pinched.3pSg the frost.NOM the garden.INESSIVE

“The buds were damaged by the frost (in the garden).”

Topic(s) may be immediately preceded, followed, or (given multiple topics) separated by a sentence adverbial, distinguishing the topic position from the focus position, which may *not* be immediately followed by sentence adverbials (É. Kiss, 2008). This difference is exemplified in (4) below.

[4] Sentence adverbials following a topic versus a focus

a. Topic + sentence adverbial

A: *Did Eeyore make it to the train?*

B: *Igen, [TOPIC Füles] [COMMENT szerencsére el- érte a vonatot.]*
yes Eeyore.NOM luckily VPX- reached.3pSg the train.ACC
“Yes, Eeyore made it to the train, luckily.”

b. Focus + sentence adverbial

A: *Who made it to the train?*

B: **[FOCUS Füles] szerencsére érte el a vonatot.*
Eeyore.NOM luckily reached.3pSg VPX the train.ACC
“Eeyore made it to the train, luckily.”

Another structural difference between topics and foci is that the topic, unlike focus, is not assumed to trigger verb movement. If a verbal modifier is present in the sentence, the topic appears in a (neutral) pre-verbal position, as opposed to focus, which triggers the leftward movement of the verb over the verbal modifier (É. Kiss, 2002). See a schematic representation of this difference in (5).

[5] Schematic *Topic + Comment* word order with and without focus:

a. Topic + [COMMENT Focus + Verb + Verbal modifier + post-verbal elements]

A: “What did the frost damage?”

B: *[TOPIC A dér] [COMMENT [FOCUS a rügyeket] csípte meg]].*
the frost.NOM the buds.ACC pinched.3pSg VPX,
“The frost damaged the BUDS .”

b. *Topic + [COMMENT Verb + Verbal modifier + post-verbal elements]

A: “How are the buds in the garden?”

B: **[TOPIC A rügyeket] [COMMENT csípte meg a dér (a kertben)].*
the buds.ACC pinched.3pSg VPX the frost.NOM the garden.INESSIVE
“The buds were damaged by the frost (in the garden).”

Topics in Hungarian thus appear in a pre-comment position, which is syntactically distinguished from the focus position by being able to immediately precede a sentence adverbial and by not triggering verb movement.

With respect to the role of topics in a sentence, É. Kiss (2008) suggests the topic to be the logical subject of predication, and the comment to be the predicate – a suggestion compatible with that of Reinhart (1981) and Krifka (2008) introduced above. Similarly to Reinhart (1981), referring expressions are suggested as possible topics, and as such, É. Kiss (2008) argues that syntactic subjects, objects, and PPs containing a referring expression may act as topics in Hungarian. The requirement of referentiality holds true unless the topic is meant to be contrastive, in which case non-referential expressions, such as bare nominals, predicative adjectives, adverbials, infinitivals, and verbal prefixes may occur in topic position as well (see Gyuris, 2002).

Krifka (2008) and Krifka & Musan (2012) suggest that contrastive topics comprise an *aboutness* topic with an associated focus, proposing that these topics evoke alternatives (i.e. are contrastive) because they contain a focus. A number of questions arise from this proposal and the foregoing discussion. However, for the purposes of this dissertation, these questions are put aside, and the paragraphs below review some of the characteristics of (*aboutness*) *topics* and *contrastive* topics, with a focus on tendencies for their prosodic expression in Hungarian.

1.4.1.1.2 Two types of topic

1.4.1.1.2.1 Referential (aboutness) topic

Aboutness topics are referential expressions of which the comment is said to be about (Krifka, 2008; Reinhart, 1981). In Hungarian, such topics can be “(n)ames, definite noun

phrases, and specific indefinite noun phrases (or PPs subsuming such a noun phrase)”; where “specific” means that the referent exists in the domain of discourse (É. Kiss, 2008:288). In Hungarian, referential (aboutness) topics occur in the syntactic topic position introduced above. When *given*, these types of topics may be dropped. *Füles* in (4) above and *Hume* in (6/a) below are referential topics.

1.4.1.1.2.2 Contrastive topic

Contrastive topics are felicitous when at least one alternative topic is available. Krifka & Musan (2012) suggest that contrastive topics are used to “indicate that the current clause does not deliver all the information that is expected” and are thus often used in incremental responses (pp. 30). In example (6/b) below, the question already suggests the possibility of alternatives, and so the (incremental) response has two contrastive topics corresponding to each alternative. In this case, the lack of complete information in the first clause is expected given the likelihood of alternatives present in the question. In example (6/c) below, the question introduces a type of entity (*cruciferous vegetable*) of which *broccoli* in the answer is a subtype. Answering with information about *broccoli* first gives rise to the possibility that something other than what is stated for *broccoli* may apply to other types of cruciferous vegetables, and thus leaves open the possibility that more information can be expected. Example (6/d) below is a more implicit case in which the question suggests one possibility (being a bit of a coward) and the response negates the opposite (being brave), thereby implicitly suggesting the alternative, or the opposite end of the scale for non-categorical adjectives. The specificity of the generated implicature, i.e. the alternatives, likely depends on their own contextual salience. For a formal theory of contrastive topics, see Büring (2003).

In Hungarian, referential and non-referential contrastive topics occur in the same left periphery topic position as do contrastive and non-contrastive referential topics (É. Kiss, 2002). Expressions that can only occur in topic position when they are contrastive include bare nouns, quantified NPs, adjectives, adverbs, and infinitives. A contrastive example of a referential topic is given in (6/b) and of bare nouns and adjectives in (6/c, d).

[6] Examples of topics

a. Referential (aboutness) topic

A: *What did Hume unwittingly do?*

B: [TOPIC *Hume*] [COMMENT *fel-ébresztette Kantot a dogmatikus szendergéséből*].
 Hume.NOM VPX-woke.3pSg Kant.ACC the dogmatic slumber.POSS.ELATIVE
 “Hume woke Kant up from his dogmatic slumber.”

b. Contrastive topic

A: A: *Who did what unwittingly?*

B: [TOPIC *Hume*] [COMMENT *fel-ébresztette Kantot a dogmatikus szendergéséből*].
 Hume.NOM VPX-woke.3pSg Kant.ACC the dogmatic slumber.POSS.ELATIVE
 és [TOPIC *Austen*] [COMMENT *idegesítette Brontë*].
 and Austen.NOM annoyed.3pSg Brontë.ACC
 “Hume woke Kant up from his dogmatic slumber and Austen annoyed Brontë.”

c. Contrastive bare nominal

A: *Was it a good idea to serve cruciferous vegetables or did we end up wasting a lot of food?*

B: [TOPIC *Brokkolit*] *szerencsére* [COMMENT *kért a Mama*],
 broccoli.ACC luckily request.3pSg the Grandma.NOM
 a többiből mindenki evett.
 the rest.ELATIVE everybody.NOM ate.3pSg
 “Grandma asked for broccoli, luckily, and everyone had some of the rest [of the cruciferous vegetables].”

d. Contrastive adjective (adapted from Gyuris (2012))

A: *Are you a chicken?*

B: *Hát, [TOPIC bátor] sajnos nem vagyok.*

well brave sadly not be.1pSg

“Well, I am not brave, sadly” / “*Brave* I am not, sadly.”

From aboutness and contrastive topics, Krifka (2008) further distinguishes certain topic-adjacent phenomena, namely, *frame-setting* and *delimitation*, shown in (7) below. These have been argued to restrict the domain of interpretation (7/a) or specify the dimension of evaluation (7/b) (Krifka 2008, Jacobs 2001). Gyuris (2012) suggests that Hungarian has similar topic-adjacent dimensions. These topic-adjacent expressions will not be reviewed here further and will not be examined in the dissertation.

[7] Frame setting and delimitation

a. Frame setting (adapted from Jacobs, 2001)

A: *How is Peter?*

B: [FRAME *Körperlich*], *geht es Peter* [FOCUS *GUT*].

physically goes expl. Peter.dat well

“Physically, Peter is well.”

b. Delimitation (Krifka, 2008)

[DELIMITATION *An inGENious mathematician*] *he is* [FOCUS *NOT*]

1.4.1.1.3 The prosodic expression of topics in Hungarian

In addition to regularities in word order, topics may also be marked by prosody in Hungarian. É. Kiss (2008), for instance, suggests that the topic constituent appears before the first pitch accent of the comment. The topic constituent itself can bear prosodic prominence as well (Varga, 2017; Genzel et al., 2015; Mády, 2015; Gyuris & Mády, 2014), and the shape and

relative prominence of the pitch contour on the topic may be modulated by certain factors. One factor is the topic's information status; that is, whether it is given or new in the discourse. In particular, Genzel et al. (2015) find that if the topic is *given*, a rising contour is more likely over a falling one. A further factor is the type of focus that might follow the topic. For instance, Genzel et al. (2015) report fewer falling and more flat and rising contours on the topic when it is followed by a contrastive focused constituent. Besides contrastive focus, non-contrastive preverbal focus can partially affect the distribution of pitch accent types on the topic, such that the f0 peak may be lower in a rising pitch accent on a topic than on a focus, perhaps corresponding to its relatively smaller prominence (Genzel et al., 2015).

In terms of the *timing and shape of the contours*, Gyuris & Mády (2014) report late peaks to be most characteristic of proper name and bare nominal topics, and mid peaks to occur more frequently with quantified NPs. The timing of the peak (early, late, mid) refers to the relative position of the pitch target in or around the accented syllable. In terms of f0 peak height, f0 excursion, f0 slope (steepness), and peak alignment within the syllable, Gyuris & Mády (2014) do not find significant differences between contrastive and non-contrastive topics, in contrast with reports on German (Mehlhorn, 2001) and Italian (Braun, 2006).

Contrastive topics are suggested to often be produced with a rising pitch accent (Krifka & Musan, 2012). Building on the idea that contrastive topics contain a focus, Krifka & Musan (2012) argue that it is the focus in contrastive topics that is marked by the rising accent, and this accent is not the most prominent one in the sentence (which instead occurs somewhere in the *comment*). The observation of rising accents is true for English and Greek: speakers produce rising bitonal pitch accents, either L+H* for contrastive topic (Hedberg, 2003; Stavropoulou & Baltazani, 2021) or L*+H for corrective topic (Stavropoulou & Baltazani, 2021). Neapolitan

Italian, on the other hand, marks contrastive topics by phrasing instead of pitch accentuation, with contrastive topics constituting their own prosodic phrase (D’Imperio & Cangemi, 2011).

In Hungarian, it has been suggested, based on impressionistic analysis, that contrastive topics either consist of a rise followed by a pause (Szabolcsi, 1981) or of a (fall)-rise (É. Kiss, 1994). Experimental work found, however, that contrastive topics are not necessarily marked by a rising contour. Gyuris & Mády (2014) report flat contours (*plateaus*) rather than rises to be the most common. Exceptions to this are *given* bare noun topics, where rises are more frequent, and *new* (contrastive) proper name topics, where there appear to be as many rises as plateaus. The second most common contour is the rise, except in *given* (non-contrastive) proper name topics, where the second most common contour is the fall (see these results summarized in Gyuris & Mády (2014:11). Importantly, falls do not necessarily precede contrastive topics and rises do not distinguish between contrastive and non-contrastive topics. The introspective observation that a falling contour precedes a rise in contrastive topics is not supported in Gyuris & Mády (2014).

One semi-regular difference can be found in the *overall* contour of contrastive vs. non-contrastive proper name topics. Gyuris & Mády (2014) find the *tail* (the unstressed syllables following the stressed and accented syllable) of all topic contours to be most frequently a plateau, but *if* the tail contour is rising, the proper name topic is usually contrastive, whereas if the tail contour is falling, the topic is usually non-contrastive. Regarding bare nominal and quantified NP topics, the study finds no significant difference in the tail contour based on their information status. Gyuris (2012) suggests that, at least for proper name topics, it is not the corresponding pitch accent that differentiates contrastive from non-contrastive, but the contour from the proper name topic pitch accent to the pitch accent on the following focus.

The presence of mid and late peaks, as well as the possibility of a rising contour in given and contrastive topics is important to keep in mind for the purpose of examining the temporal alignment of prosodic and co-speech gestural targets between different types of prominent units. Since Gyuris & Mády (2014) suggest that it may be the tail contour that differentiates between contrastive and non-contrastive proper name (referential) topics, such topics could provide an interesting test case for whether the target of the co-speech gesture aligns with the stressed syllable, which in Hungarian is the first syllable, or with the f0 target corresponding to the rising contour, which may be reached at the end of the topic phrase (i.e. not during the first syllable in a multisyllabic expression).

This concludes the discussion of topics and their expression in Hungarian. The next subsection reviews the information structural concept of focus, focusing on its structural and prosodic expression in Hungarian.

1.4.1.2 Focus

The present subsection reviews focus with an emphasis on its expression in Hungarian (first, its structural and pragmatic dimensions, then separately its prosody).

Rooth (1992) describes the meaning of focus for English as related to “contrast within a set of alternative elements” (pp. 113). Pragmatic uses of focus, according to Krifka (2008), are to highlight information (in particular, the part of the sentence that corresponds to a WH-element in an explicit or implicit question that the sentence answers), and to confirm or correct. The example below in (8) illustrates such a question-answer focus, where the alternative set (in the sense of Rooth’s *Alternative Semantics*) consists of authors, and the focus-semantic meaning of the sentence is a set of propositions that differ in the focused element, in this case the identity of the author (i.e. {*J. K. Rowling* wrote those books of the wizard boy with glasses, *Stephen King*

wrote those books of the wizard boy with glasses, *Roald Dahl* wrote those books of the wizard boy with glasses, ...,}).

[8] Focus

A: [*Who*] wrote those books of the wizard boy with glasses?

B: [_{FOCUS} *J. K. Rowling*] did.

1.4.1.2.1 The syntactic focus position

Focusing on Hungarian, parallel with her treatment of topics, É. Kiss (2008) argues that there exists a focus position on the left-periphery (SpecFP, the specifier of a functional projection). Further, É. Kiss (2008) notes that focus attracts the verb to move (across the verbal modifier, when present) into a position adjacent to focus. The examples in (9) below make use of the question-answer format to illustrate focus and demonstrate that only the response in (9/a), in which the focused constituent is immediately followed by the verb, is an acceptable response to a WH-question. (9/b) illustrates that even if it is only a verbal prefix (VPX) that precedes the verb thereby intervening between the focus-verb sequence, the sentence is not an acceptable response. (9/c) shows that further constituents may also not precede the verb and immediately follow the focused constituent. Finally, (9/d) shows that this is the case even if the verb in fact moves across the verbal prefix. The sentences in (9/b-d) are grammatical, but they answer different questions. Coincidentally, (9/d) also contains a preverbal focus, answering the question “When did David see Oppenheimer in the movie theaters?”.

[9] Focus position

a. A: *What did Dávid see at the movie theater in August?*

B: [TOPIC *Dávid*] [COMMENT [FOC *az Oppenheimert nézte meg Augusztusban*].
 Dávid.NOM the Oppenheimer.ACC watched.3pSg VPX August.INESSIVE
 “David saw OPPENHEIMER in August.”

b. A: *What did Dávid see at the movie theater in August?*

B: *[TOPIC *Dávid*] [COMMENT [FOC *az Oppenheimert*] *meg- nézte Augusztusban*].
 Dávid.NOM the Oppenheimer.ACC VPX-watched.3pSg August.INESSIVE
 “David saw OPPENHEIMER in August.”

c. A: *What did Dávid see at the movie theater in August?*

B: *[TOPIC *Dávid*] [COMMENT [FOC *az Oppenheimert*] *Augusztusban meg-nézte*].
 Dávid.NOM the Oppenheimer.ACC August.INESSIVE VPX-watched.3pSg
 “David saw OPPENHEIMER in August.”

d. A: *What did Dávid see at the movie theater in August?*

B: *[TOPIC *Dávid*] [COMMENT [FOC *az Oppenheimert*] *Augusztusban nézte meg*].
 Dávid.NOM the Oppenheimer.ACC August.INESSIVE watched.3pSg VPX
 “David saw OPPENHEIMER in August.”

There appears thus to be a preverbal focus position in Hungarian, which, when filled, appears to trigger verb movement systematically. This is also the case when certain overt focus operators (particles that affect the interpretation of the focused unit) are present, such as *csak* “only” and *még* “additionally”. Conversely, with the focus operators *még ... is* “even”, and *is* “too”, verb movement appears to not be triggered. The differences are illustrated in (10-11).

[10] Focus operators in Hungarian - movement

a. *csak* (‘only’):

A: *What did the frost damage?*

B: *Csak* [FOCUS *a rügyeket*] *csípte* *meg*.
 only the buds.ACC pinched.3pSg VPX
 “It only damaged the BUDS.”

B: **Csak* [FOCUS *a rügyeket*] *meg-csípte*.
 only the buds.ACC VPX-pinched.3pSg
 “It only damaged the BUDS.”

b. még (‘also’, ‘additionally’)

A: *The frost damaged the grass and the leaves, didn’t it. What else did it damage?*

B: *Még* [FOCUS *a rügyeket*] *csípte* *meg*.
 additionally the buds.ACC pinched.3pSg VPX
 “It also damaged the BUDS.”

B: *?*Még* [FOCUS *a rügyeket*] *meg-csípte* .
 additionally the buds.ACC VPX-pinched.3pSg
 “It also damaged the BUDS.”

[11] Focus operators in Hungarian - no movement

a. még ... is (‘even’)

A: *The frost damaged all kinds of things, did it not?*

B: *Még* [FOCUS *a rügyeket*] *is meg-csípte*.
 even the buds.ACC too VPX-pinched.3pSg
 “It even damaged the BUDS.”

B: **Még* [FOCUS *a rügyeket*] *is csípte* *meg*.
 even the buds.ACC too pinched.3pSg VPX
 “It even damaged the BUDS.”

b. is (‘too’, ‘also’)

A: *The frost only damaged the grass, didn’t it?*

B: [FOCUS *A rügyeket*] is *meg-csípte*.
 the buds.ACC too VPX-pinched.3pSg
 “It damaged the BUDS too.”

B: *[FOCUS *A rügyeket*] is *csípte* *meg*.
 the buds.ACC too pinched.3pSg VPX
 “It damaged the BUDS too.”

Two other instances in which focus is not necessarily preverbal are the case of double foci with one focused element occurring in the canonical preverbal focus position, and the other following the verb, and the case of listing items. These are noted here (a) to present a more complete picture of the intricacies of the syntactic focus position and (b) to caution against too broad generalizations about Hungarian focus. This line of thought will not be pursued further in the dissertation, which is mainly concerned with focus cases in which the unit is clearly preverbal.

1.4.1.2.2 Types of focus

The previous subsection introduced the syntactic focus position. In the above examples, focus is the part of a sentence that provides information to an implicit or explicit question (very broadly, examined through a *Question Under Discussion* (QUD) approach (Roberts, 1996, 2012)). Although a lot can be generalized over and subsumed under this approach (possibly all types of pragmatically defined foci), the following examples are meant to differentiate between types of focus in terms of their pragmatic import: informational, exhaustive, contrastive, corrective, and confirmation focus. The subsection will not review further focus types, such as second-occurrence focus and verum/falsum focus.

1.4.1.2.2.1 Informational and exhaustive focus

In Hungarian, movement into the syntactic focus position is argued to trigger an *exhaustive* interpretation of the focused element, indicating that, given some context, only with that specific focused element is the proposition true (É. Kiss, 2008). In (12) below without the topic constituent, no other movie outside of *Oppenheimer* is understood to have been seen by *Dávid* at the movie theater in August. When the speaker who provides the answer is aware that *Dávid* only watched two movies, placing only one (*Oppenheimer*) in focus position is odd (13/a) and suggests he only watched that one movie. Placing two in focus position via coordination (13/b) is natural, since it provides a complete answer to the question. Placing part of the response in a postverbal position is also natural (13/c). This latter example contrasts with exhaustive focus in that it adds new information to the discourse, but is not associated with the interpretation that *Dávid* only watched *Oppenheimer* and no other movie. Such a focus is said to be *informational* (É. Kiss, 2008). In fact, É. Kiss (2002, 2008) argues that in Hungarian, focus in a postverbal position is generally informational and non-exhaustive.

[12]

A: *What did Dávid see at the movie theater in August?*

B: [COMMENT [FOCUS *az Oppenheimert*] *nézte meg Augusztusban*].
the Oppenheimer.ACC watched.3pSg VPX August.INESSIVE
“(He) saw OPPENHEIMER in August.”

[13] Context: Dávid only saw *Oppenheimer* and *Barbie* at the movie theater in August.

A: *What did Dávid see at the movie theater in August?*

a.

B: *[COMMENT [FOCUS *Az Oppenheimert*] *nézte* *meg Augusztusban*].
the Oppenheimer.ACC watched.3pSg VPX August.INESSIVE
“(He) saw OPPENHEIMER in August.”

b.

B: [COMMENT [FOCUS *Az Oppenheimert és a Barbie-t*] *nézte* *meg Augusztusban*].
the Oppenheimer.ACC and the Barbie.ACC watched.3pSg VPX August.INESSIVE
“(He) saw OPPENHEIMER and BARBIE in August.”

c.

B: *Hát...* [COMMENT *meg-nézte* [FOCUS *az Oppenheimert*].
Well VPX-watched.3pSg the Oppenheimer.ACC
“Well, (he) saw OPPENHEIMER in August.”

Although the intuition of preverbal exhaustivity has been suggested to be rather strong in Hungarian, recent empirical evidence suggests a need for revision of the strong binary division in the exhaustivity dimension between preverbal and postverbal focus, as given the right context, the exhaustive interpretation of preverbal focus may be cancelled (see e.g. Gerőcs, Babarczy, & Surányi, 2014; Destruel et al., 2015; Káldi, Babarczy, & Bende-Farkas, 2017; Káldi and Babarczy, 2018).

1.4.1.2.2.2 Contrastive focus

While focus is in general understood as implying alternatives, the nature of the alternative set can change or add to the pragmatic functions of focus. An explicit (14/a) or definite, but not explicitly stated (14/b) alternative set can highlight the contrast between the focused element and the alternatives, as illustrated by the example below. This type of focus is referred to as *contrastive*. It has been suggested that *contrast* may be gradient (Paoli, 2009; Molnár, 2006), for

a discussion, see Repp (2016), and so there are further divisions and/or possible degrees of contrast having to do with, for instance, whether the alternative set is explicit or implicit in the preceding discourse, the size of the alternative set, or the pragmatic purpose of the contrast. The examples below illustrate that contrastive focus occurs in the preverbal focus position.

[14] Contrastive focus

a. Contrastive focus with an explicit alternative set

A: *Did Peter finish reading a book or a newspaper?*

B: *Péter* [FOCUS *egy könyvet*] *fejezett be, nem egy újságot.*
Péter.NOM a book.ACC finished.3pSg VPX not a newspaper.ACC
 “Péter finished a BOOK, not a newspaper.”

B: **Péter be-fejezett* [FOCUS *egy könyvet*], *nem egy újságot.*
Péter.NOM VPX-finished.3pSg a book.ACC not a newspaper.ACC
 “Péter finished a BOOK, not a newspaper.”

b. Contrastive focus with a definite but not explicitly stated alternative set

A: *Which Harry Potter book is your favorite?*

B: [FOCUS *Az ötödiket*] *szeretem leginkább.*
the fifth.ACC like.1pSg most
 “I like the fifth one the most.”

B: **Leginkább szeretem* [FOCUS *az ötödiket*].
most like.1pSg the fifth.ACC
 “I like the fifth one the most.”

1.4.1.2.2.3 Corrective focus

A further, more specific, type of contrast is correction relative to information asserted (15/a) or presupposed (15/b) in previous discourse. In both of the examples given below, the alternative set has one member, but this need not be the case. Importantly, in terms of discourse

functions, the corrective focus in B's response overrides the information entered into the common ground by A (15/a), or it sabotages its entry by asserting another alternative and excluding the one offered by A (15/b). This type of focus is referred to as *corrective*. The examples below illustrate that it occurs in the preverbal focus position naturally (although it is acceptable post-verbally if the corrected unit is part of a quote).

[15] Corrective focus

a. Correcting an assertion

A: *Péter be-fejezte a könyvet.*

Péter.NOM VPX-finished.3pSg the book.ACC

"Péter finished the book."

B: *Nem, [FOCUS az újságot] fejezte be.*

no the newspaper.ACC finished.3pSg VPX

"No, he finished the NEWSPAPER."

b. Correcting a presupposition

A: *Be-fejezte Péter a könyvet?*

VPX-finished.3pSg Péter.NOM the book.ACC

"Did Péter finish the book?"

B: *Nem, [FOCUS az újságot] fejezte be.*

no the newspaper.ACC finished.3pSg vpx

"No, he finished the newspaper."

1.4.1.2.2.4 Confirmation / Agreement focus

A further focus type is what we may call *confirmation* or *agreement focus*. Here, the constituent in question is *given* and marked for focus at least morpho-syntactically or syntactically both in the first and second occurrences. The sentence containing the focused constituent has the pragmatic function of agreement or confirmation. In (16/a) below, the

focused constituent in B's response was set up to be focused in A's question (in both, focus is indicated by capital letters in the English translation). Given other contextually salient contestants (such as *Barbi*, *Katica*, *Szilvi*, ...), semantically, the proposition *Gitti won today* is presented in A's question as the 'preferred alternative', with the others being {*Barbi won today*, *Katica won today*, *Szilvi won today*, ...,}, where the focused constituent (*Gitti*) gives rise to the focus semantic meaning of the yes/no question. For illustration, (16/b) is given as a different polar question in which *Gitti* is not focused.

[16] Confirmation focus

a. Focus on *Gitti* in the question

- A: *Gitti nyerte meg a meccset?*
 Gitti.NOM won.3pSg VPX the game
 "Did *GITTI* win the game?" / "Was it *GITTI* that won the game?"
- B: (*Igen*,) [_{FOCUS} *Gitti*] *nyerte meg (a meccset).*
 yes Gitti.NOM won.3pSg VPX the game
 (Yes,) *GITTI* won (the game).

b. No focus on *Gitti* in the question

- A: *Gitti meg-nyerte a meccset?*
 Gitti.NOM VPX-won the game.ACC
 "Did Gitti win the game?"
- B: (*Igen*,) *Gitti meg-nyerte (a meccset).*
 yes Gitti. NOM VPX-won the game.ACC
 (Yes,) Gitti won (the game).

This subsection reviewed some of the main syntactic and/or pragmatic types of focus with a focus on their expression in Hungarian. Importantly, aside from *information* focus, the focus types discussed (exhaustive, contrastive, corrective, confirmation) all tend to occur in the

canonical preverbal focus position. The next section reviews research on the prosody of focus in Hungarian.

1.4.1.2.3 The prosodic marking of focus in Hungarian

As shown above in the present section, Hungarian marks focus syntactically; however, there are certain tendencies in additional prosodic marking as well. The present subsection reviews these tendencies in order to see what may be expected in speech and co-speech gesture alignment. As *section 1.3* has shown, much of the literature examining the alignment of speech and co-speech gestures which report peak-to-peak alignment involves languages in which focus is primarily marked by intonation (in particular, pitch accentuation), with very few studies on languages where focus is primarily signaled via word order.

As focus is marked syntactically in Hungarian, researchers have stipulated that prosody may have a smaller role to play in marking focus, but found some empirical evidence to the contrary (e.g. Mády, 2015; Genzel et al., 2015; Mády, 2012). In fact, even theoretical work appeals to intuitions about prosody related to focus status: Szendrői (2003) argues that focus movement into comment-initial (immediately pre-verbal) position is motivated by prominence-seeking; i.e. the element to be focused takes the first position in the comment in order to receive the greatest prosodic prominence (Hungarian is left-headed in its prosodic structure, see Varga (2002)). Whether it is correct in the *movement-motivating* role of prosodic prominence or not, Szendrői's (2003) account stipulates that preverbal focus is in fact marked prosodically. In the following, the focus-related correlates of prosodic prominence are reviewed: pitch accentuation on the focused element and the potential deaccentuation of the background, longer vowel duration, pause insertion before the prominent unit, increased intensity, and changes in the properties of vowels.

1.4.1.2.3.1 Pitch accentuation and deaccentuation

The pitch-accentuation of focused constituents and the deaccentuation of background units is a common strategy in focus marking, attested in various languages; for a recent review of these, see Kügler & Calhoun (2020). In Hungarian, in *broad focus* sentence contexts, most content words are pitch accented. Mády & Kleber (2010) report content words to be deaccented 21% of the time overall, and 3% of the time utterance-initially in read speech. In *narrow focus* contexts, previous theoretical and introspective work expected deaccentuation of the background (Varga, 2002; Kálmán & Nádasdy, 1994). Experimental work finds pre-verbal focus to be indeed usually pitch-accented, but background elements to not be necessarily de-accented (Genzel et al., 2015; Mády, 2015). Accordingly, Mády (2015) suggests that the deaccentuation of the background is optional rather than obligatory even in narrow and contrastive focus contexts, although she does find it to be more likely to occur given a narrow or contrastive focus (50%) than in a broad focus condition (32%). Consistent with this, Genzel et al. (2015) find that the deaccentuation of the background is modulated in part by the background's information status, that is, *given* background elements are more likely to be deaccented than *new* ones are. Additionally, deaccentuation has been found to be more likely in the post-focal background region than in pre-focal background elements (Genzel et al., 2015). Given the finding that the background is deaccented in narrow and contrastive focus conditions only 50% of the time, Mády (2015) concludes that marking focus via pitch accentuation against a deaccented background is optional and not systematically done in Hungarian.

In an IP that stretches over a focused item, the nuclear pitch accent falls on the first content word of the preverbal focus phrase. Further exploring the role of prosody in Hungarian focus marking, Langer & Kügler (2021, 2023) find that this is so even if the focus domain is

smaller within the phrase in focus position; in particular, in complex NPs with narrow focus not on the first word or the entire phrase (18/a) but on a subsequent word within the focus phrase (18/b). Their finding is consistent with the idea proposed by É. Kiss (2002) that the first constituent of the *predicate* must bear the heaviest pitch accent. Additionally, Langer & Kügler (2021, 2023) report finding a second, downstepped pitch accent, the placement of which largely co-varies with the focus domain within the focus phrase. A notable exception to this pattern, as pointed out by the authors, is in the lack of significant prosodic difference found between two different focus domains corresponding to different words in the modification of the NP; i.e., focus on the adjective vs. on an argument of the adjective is not distinguished in prosody. The authors interpret the results as prosody playing a disambiguating role in distinguishing focus domains within complex phrases in the syntactic focus position; this function, however, is not absolute. Further exploring prosodic patterns within larger focus phrases (as in 18/c), Kügler et al. (2026) examine the elements preceding the narrowly focused one and find that within such multi-word focus phrases, the leftmost content word (*kicsi*) receives the highest prominence, and following elements (*piros*) leading up to narrow focus (*book*) have reduced prominence. The authors view these findings as the prosodic attenuation of pre-(narrow)-focal elements contained in the larger focus phrase, so as to increase (relatively speaking) the prominence of the narrow-focused element.

[18] Complex NP/DP focus

a. Whole NP/DP focus domain

A: *What did you find?*

B: [FOCUS PHRASE [FOCUS *Egy piros könyvet*]] *találtam*.

a red book.ACC find.1pSg

“I found *a red book*.”

b. Smaller NP/DP focus domain

A: *What red object did you find?*

B: [FOCUS PHRASE *Egy piros* [FOCUS *könyvet*]] *találtam.*

a red book.ACC find.1pSg

“I found a red BOOK.”

c. Longer NP/DP focus domain

A: *What small red object did you find?*

B: [FOCUS PHRASE *Egy kicsi piros* [FOCUS *könyvet*]] *találtam.*

a small red book.ACC find.1pSg

“I found a small red BOOK.”

Regarding *pitch accent type* in Hungarian, Mády (2015) does not find a significant difference based on focus type (broad, narrow, and contrastive). The author reports different falling pitch accents to be the most frequent (mainly H+L*, and a small percentage of H*+L and L*), the distribution of which also does not vary by focus type. Similarly, Genzel et al. (2015) finds a tendency for falling pitch accents on focused items. Background elements, when pitch accented, are reported in Mády (2015) to also have a H+L* accent. For the purposes of the dissertation, it is noteworthy that the authors of both studies find a tendency for a *low* pitch accent target in focus rather than a high one, contrary to observations in many other languages in which L+H* and H* pitch accents tend to mark focus (see Kügler & Calhoun, 2020). In this respect, focus and topic prominence appear to diverge with a tendency for rising contours for topic (with a high pitch target) and falling and low contours for focus (with a low pitch target). Two questions arising from these focus marking tendencies are (a) how to capture alignment in expressions with a low pitch accent target and (b) whether a co-speech gesture target aligns more closely or in a more stable manner with the prominent low target of the pitch accent or with the f0 maxima.

Regarding the *shape of the focus pitch accent*, findings vary. In some languages, the relative height of the f0 maxima can signal differences in focus type; in particular, contrastive vs. non-contrastive (see the discussion in Kügler & Calhoun (2020)). In Hungarian, while Genzel et al. (2015) report higher f0 maxima and a steeper fall on the focused element (broad < narrow < contrastive), Markó et al. (2018) do not find a focus effect on f0. However, even though H+L*, L*, and H*+L most frequently occurred on the focused unit, listeners in Mády's (2015) perception task do associate a higher f0 maxima with narrow focus – lending support to the idea that relative f0 height can signal focus in Hungarian. Given the fact that the L target is associated with the prominent syllable in these pitch accents (except in H*+L), that f0 height should be perceptually important may signal that the relative difference between the H and L targets in the bitonal pitch accents is an important factor in focus marking.

In terms of the *alignment of the f0 target* within the accented syllable, reports again vary. On the one hand, Mády (2015) finds *later* alignment in focus (that is, the pitch accent target occurs later in the syllable), while Markó et al. (2019) report an *earlier* peak in (utterance-initial) focus than in topic. (Note however that Markó et al. (2019) do not report on the *type* of pitch accents in their data, rendering it difficult to compare the two studies. As stated above, Mády (2015) finds falling contours to be the most frequent.) It appears then that there is no consensus on the alignment of the peak within the prominent syllable, which is important to note as a potential source of variability in the alignment of speech with co-speech gestures, provided that the co-speech gesture coordinates with the pitch accent.

1.4.1.2.3.2 Phrasing: pauses and boundary-related lengthening

In addition to marking focus syntactically and by pitch accentuation, another strategy employed by some languages is phrasing; that is, placing a boundary before, after, or around the

focused phrase (see Kügler & Calhoun (2020) for a review of focus marking in Korean and Chichewa). Hungarian might also employ phrasing for this purpose. Although Mády et al. (2016) finds Hungarian speakers to pause more frequently in general than do a comparison group of German speakers, pauses could occur for multiple reasons and pragmatically prominent units are indeed found to often be preceded by a pause in Hungarian (Mády et al., 2016; Mády & Kleber, 2010). In particular, Mády et al. (2016) report speakers to pause more often and for longer periods before prominent (pitch accented) words than before non-prominent ones. Consequently, the authors suggest that stronger prosodic boundaries precede more prominent units.

1.4.1.2.3.3 Vowels under focus prominence

Regarding vowel quality under focus, formant variability has been shown to be smaller (Markó et al., 2018). In comparing utterance-initial vowels in two syntactic positions (either pre-focus topic position, or preverbal focus position), Markó et al. (2019) similarly find smaller variability in the F1xF2 space in focus position, but do not find a more peripheral articulation of the vowels. Potentially due to vowel length being phonemic in the language, lengthening under stress is argued to not be a phonetic correlate of *stress* (see Vogel, Athanasopoulou, & Pincus (2015) and White & Mády (2008), but see also Szalontai et al (2016) for the opposite). When the stressed syllable is *prosodically prominent* as well, however, vowel duration is longer (Markó et al., 2019; Mády & Kleber, 2010, but see Vogel, Athanasopoulou, & Pincus (2015) for nuanced results related to phonological vowel length). Intensity measured in the vowel is affected by accent (Szalontai et al., 2016).

This concludes the review of the prosodic marking of focus in Hungarian. As can be seen, there are a few tendencies in prosodic marking, which include the frequent accentuation of the prominent phase (often with a low pitch accent target) and de-accentuation of the

background, and frequent pausing preceding the focus phrase. The phonetic effects of focus prominence on the vowel in the prominent syllable include smaller variability in its articulation, lengthening, and higher intensity. Importantly, in contrast to other languages, there is evidence in Langer & Kügler (2021, 2023) that the highest prominence in the Hungarian sentence is not necessarily the (narrowly) focused unit, but the left-edge of the focused phrase, which coincides with the left edge of the comment or predicate.

1.4.1.3 Information status

The last information structural notion to be reviewed, rather briefly, is *givenness*, which refers to the information status of an expression. *Given* vs. *new* have been defined in a number of ways, but generally they are conceptualized as the presence vs. absence of an expression in the communicative context of a sentence. Semantically, the denotation of an expression can be *given* or *new* in the discourse; the former term is applied if the denotation is understood to be present in the common ground, and the latter when it is not (Krifka & Musan, 2012). Götze et al. (2007) also include *accessible* as a possible information status, defined for a discourse referent which is accessible via a “relation to a referent in the previous discourse, in the situative context, or the assumed world knowledge of the hearer” (Götze et al., 2007:156). Examples are given below in (19).

[19] Information status

a. New and given

[_{NEW} *Bogyó*] *has finally caught the ball.* [_{GIVEN} *She*] *seems pleased with herself.*

b. Accessible

Bence went to [_{NEW} *the zoo*]. *He was most excited to see* [_{ACCESSIBLE} *the lions*].

That an expression is given may be indicated by a number of linguistic means, (see Gyuris, 2012; Krifka & Musan, 2012), among them pronominalization, deaccentuation, use of definite descriptions, and deletion.

A peculiarity occurs with a mismatch between *discourse given* in the sense introduced in this subsection and *given as focused*. The following example adapted from Schwarzschild (1999) in (20a) illustrates this point.

[20] To be given or not to be (given)

a. English (adapted from Schwarzschild, 1999)

A: *Who did John's mother praise?*

B: *She praised* [_{FOCUS} *HIM*].

b. Hungarian (adapted from Gyuris, 2012)

A: [_{FOCUS} *János*] *főzte meg a vacsorát?*
 János.NOM cooked.3pSg VPX the dinner.ACC
 “Did János cook dinner?”

B: *Igen*, [_{FOCUS} *Ő*].
 Yes personal-pronoun.NOM
 “*Yes, he did.*”

John is discourse *given* after A's question as evidenced by the pronominalization in B's answer; therefore, any subsequent reference made to John need not include pitch accentuation in English. However, the pronoun *him* (coreferent with *John*, therefore *given*), in B's answer is marked by a pitch accent. A potential reason behind this is that *him* is not *given as focused*, rather, it provides the information sought in the WH-question. For a formal account, see Schwarzschild (1999), where the focus marking of the pronoun is instead necessitated roughly by Schwarzschild's proposed definition of givenness and its relation to the semantic effect of focus

marking. A similar example in Hungarian is given in (20/b), where the *given* proper name is pronominalized as well, instead of being completely elided, as it normally would be if it was not for the pitch accent it receives either due to being focused or due to being the first constituent of the (otherwise elided) comment. The distinction made here is only meant to highlight the importance of considering information status separately from focus (and topic, c.f. Krifka, 2008) – although this is not uncontroversial (see Schwarzschild (1999) for example).

1.4.1.3.1 Prosodic marking of information status in Hungarian

Prosodic accentuation marks newness, and – as mentioned above – deaccentuation is one of the possible markers of givenness. Although Gyuris (2012) suggests that deletion is preferred over deaccentuation in Hungarian, in experimental conditions, where speakers are required to produce *given* parts of a sentence as well, deaccentuation is a likely strategy (e.g. Genzel et al., 2015). In fact, as discussed in the subsection of focus, the information status of expressions (in particular, pre- and post-verbal background ones) modulates their accentedness (Genzel et al., 2015).

As evidenced by (20b) above, (de)accentuation, however, is constrained by sentential requirements beyond information status, as it is argued that every sentence in Hungarian must have at least one nuclear pitch accent on the first constituent of a predicate (or comment) (see Gyuris, 2012). In case only a *given* expression is available in this position, it is not deaccented; instead, it must (É. Kiss, 1981) or may (Varga, 1983) bear the nuclear pitch accent (see Gyuris, 2012).

This concludes the discussion of *givenness*. The next subsection reviews the literature on the co-occurrence of information structurally prominent and non-prominent expressions and co-speech gestures.

1.4.2 Co-speech gestural marking of information structure

As discussed in **section 1.2**, McNeill (1992) introduced the so-called *synchrony rules*. In looking at the prosody-gesture connection, **section 1.3** reviewed the relevant literature that can be construed as corresponding to the *phonological* synchrony rule. The present subsection turns to the co-speech gestural marking of information structure, which broadly corresponds to the *semantic* and *pragmatic* synchrony rules; that is, to the meaning and function related reasons behind speech and co-speech gesture coordination. *Semantic synchrony* can take many forms: co-speech gestures may be completely redundant and express the same concept as their lexical affiliate, they may complement each other, or they may be anywhere on this continuum (Bergmann et al., 2011). Examining the relationship between phonological (temporal) and semantic synchrony, Bergmann et al. (2011) find that co-speech gestures align more closely with expressions the more related they are semantically. The *pragmatic synchrony rule* predicts that speech and co-speech gestures, when produced together, have the same pragmatic function (McNeill, 1992, 2005).

Given the sometimes blurry line between semantics and pragmatics, what meaning relationship exactly counts as semantic vs. pragmatic synchrony might depend on one's theory of meaning. Supposing that co-speech gestures semantically and pragmatically align with speech at a conceptual level, what (semantic/pragmatic) meaning their lexical affiliate has, the co-speech gesture will have also. Being bound by context, one may wish to discuss the regularities found in information structural units and co-speech gestures under *pragmatic* synchrony; however, remaining agnostic as to the category of meaning *all* of information structure has (or rather, leaving room for the possibility that certain aspects of IS are context-dependent to different

degrees or in different ways), I will discuss the co-speech gestural marking of information structure rather independently of the division of the synchrony rules.

Since McNeill's (1992, 2005) conceptualization of semantic and pragmatic synchrony, research into the alignment of co-speech gestures with semantically-pragmatically prominent and non-prominent units has advanced. It may be beneficial to distinguish two lines of inquiry within this area: (1) research on the tendency of co-speech gestures to co-occur with semantically-pragmatically prominent units (i.e., how frequently a gesture is anchored to e.g. focus or topic units), and (2) research on the alignment patterns when co-speech gestures do occur on semantically-pragmatically prominent units (i.e., how closely or tightly they are aligned). A third area of study is how information structure relates to gestural realization. The section turns to the discussion of the three areas identified. With respect to the co-speech gestural marking of information structure in Hungarian, there are no studies reporting on this matter at the time of writing (with the exception of a pilot analysis related to *Chapter 3* of the present dissertation, reporting on a single speaker's production in a conference proceedings paper; see Tatár, Keshet, & Krivokapić, 2026).

1.4.2.1 Focus and topic

Studies find certain tendencies in the co-occurrence of co-speech gesture with topic and focus expressions, in their temporal alignment, and in the phonetic realization of co-speech gestures co-occurring with pragmatically prominent speech units (Carignan et al., 2024; Kügler & Gregori, 2023; Türk & Calhoun, 2024; Türk, 2020; Fung & Mok, 2018; Cruz et al., 2015; Kim et al., 2014; Loehr, 2012; Ebert et al., 2011).

1.4.2.1.1 *Co-occurrence*

A number of studies find co-speech gestures to co-occur with pragmatically prominent constituents. Loehr (2012) discusses tendencies related to *pragmatic* synchrony indirectly through the co-occurrence of gestures with certain intonational meanings in English. The author finds speakers to typically mark *focus* by a pointing gesture with an emphatic beat, *emphasis* by a beat gesture, and *contrast* by producing mirrored co-speech gestures; i.e. with *gesture A* and its mirror image *gesture B* representing contrast. Kim et al. (2014) report simultaneous head and eyebrow movements to occur more frequently on narrow focused targets relative to broad focus targets. Likewise, Cruz et al. (2015) find that pragmatic meaning has an effect on the occurrence of facial gestures: a nuclear pitch accent in a broad focus sentence is often accompanied by an up-down head movement, while in a narrow focus sentence, an additional gesture, most frequently an eyebrow movement, also occurs. Further, examining the co-occurrence of gestural phases with topic, focus, and background, Türk (2020) observes that co-speech gestures co-occur with focus and topic more than with background. Türk also notes a tendency of co-occurrence with focus rather than with topic and finds co-speech gesture phases to rarely co-occur with background elements. The author attributes these patterns to the informational import of topic and focus relative to background: the former have “active information structural functions”, which might explain their being accompanied by co-speech gestures (McNeill, 1992), as opposed to the passive background that does not add new information to the common ground. Interestingly, and in contrast with Türk’s (2020) findings, Kügler & Gregori (2023), examining German, do observe a relatively large proportion of iconic gestures to co-occur with background elements as well. Türk & Calhoun (2024) find a tendency for iconic and metaphoric gestures to occur with foci, and deictics with topics. Finally, Gregori et al. (2024) find the number of co-

speech gestures to increase with an increase in the number of pragmatic functions a focus type has (namely, *background* < *information focus* < *contrastive focus* [added contrast] < *corrective focus* [added contrast and correction]).

1.4.2.1.2 Alignment

Studies report co-speech gestures to align with pragmatically prominent units of speech. Ebert et al. (2011) report co-speech gestures to behave differently in contrastive versus information foci conditions in German, with systematicity under information focus, but not under contrastive focus. In particular, the authors find that co-speech gestures and information foci begin within less than a second of each other, with the gesture starting earlier with a systematic lag, which the authors consider to signal *onset alignment* under information focus. In contrastive focus conditions, however, the authors do not observe a clear alignment. In addition to these findings, the authors observe that co-speech gesture strokes tend to align with intonation peaks; however, they do not report on alignment as a function of focus type. Examining head and eyebrow movement under focus, Kim et al. (2014) find that when the two gestures are both produced in correspondence with the same target focus region, eyebrow movements tend to precede and head movements tend to follow the target. Probing the matter of alignment in Hong Kong Cantonese, a non-stress tone language, Fung & Mok (2018) find that the apex of pointing gestures tends to be produced within the focused word, aligning with the first syllable and the authors do not find *f0* to play a significant role in alignment. They suggest that it is the focused word that anchors the co-speech gesture. Türk & Calhoun (2024) report focus and topic units to co-occur with strokes and apexes. Finally, Kügler & Gregori (2023, 2026), examining German, find a closer phonetic alignment of co-speech gesture apexes and pitch accent peaks under focus relative to background, which the authors attribute to hyperarticulation due to focus, extending

the concept of Hypo- and Hyperarticulation (Lindblom, 1990) from speech articulation to co-speech gestures.

1.4.2.1.3 Phonetic characteristics

Examining different focus types, studies find co-speech gestures to be greater in magnitude, longer, and to have greater perceived prominence under greater pragmatic prominence. Kim et al. (2014), examining eyebrow and head movements in Australian English in broad and narrow focus conditions, find that in narrow focus, eyebrow movement is larger and peak velocity is higher at the target region than in preceding and following regions. Regarding head movements, the authors find mean displacement to be greater in narrow than in broad focus, but peak velocity to only be higher in the target and post-target regions in narrow focus (relative to pre-target), but not in the broad focus condition. Rusiewicz et al. (2014) report co-speech gestures to be longer under a contrastive pitch accent and Krivokapić et al. (2017) report the lengthening of pointing gestures under focus-related prominence as well. Looking into head movement magnitude and alignment with f0 under contrastive and corrective focus conditions, Carignan et al. (2024) find evidence that speakers of French use two different head nod strategies; one corresponding to greater range, longer duration, and less kinematic stiffness than the other. The difference between the two strategies is modulated by focus type interacting with the prosodic prominence of the focus types: both head nod strategies are used in both corrective and contrastive focus, but the magnitude of the difference between their characteristics is significant only in the corrective focus condition. The authors propose that prosodic prominence alone does not motivate the different strategies; rather, the difference between the strategies is magnified under a more “prominent” focus (understood as corrective focus with greater prosodic prominence). Finally, Gregori et al. (2024) find perceived co-speech gesture prominence to

increase with an increase in the number of pragmatic functions of a focus type (namely, *background < information focus < contrastive focus < corrective focus*).

Before moving on to information status, an idea concerning the alignment patterns in Turkish and in Mandarin Chinese proposed in McNeill (2005) is worth mentioning for its potential relevance to Hungarian. First, regarding Turkish, McNeill observes that “the gestures appeared to correspond to the spoken phrases that immediately *followed* the gesture, not the phrases with which they were synchronous” (2005:138). McNeill suggests that a mechanism he terms *self-mimicry* is responsible for this pattern: a co-speech gesture is first performed and only afterwards is it verbalized. McNeill (2005) suspects self-mimicry to be unique to Turkish, but allows for the possibility of it surfacing in other languages in which word order is free or pragmatically motivated. As there has been no prior study on co-speech gestural alignment in Hungarian, another language in which discursive needs modulate word order, the possibility of observing this pattern should be noted. Second, in Mandarin Chinese, a somewhat similar phenomenon is observed: a gesture and its lexical affiliate appear to express the same concept, while also being temporally disjointed, with the co-speech gesture, in its entirety, preceding the verbal unit. McNeill (2005) suggests that *topic-prominence* may have a role in this pattern (2005:203), observing that the co-speech gesture and the spoken phrase jointly create a framing topic. With Hungarian being topic (and focus) prominent, once again temporally offset gesture-speech sequences are a possibility.

1.4.2.2 Information status

Information status also plays a role in co-speech gesture occurrence and articulation. In general, newly introduced referents are more often accompanied by a co-speech gesture than

given referents (unless they are re-introduced with a full NP); for reviews, see Rohrer (2022) and Holler et al. (2022).

1.4.2.2.1 Co-occurrence patterns

Regarding co-occurrence patterns, Loehr (2012) finds speakers to often introduce a referent with an abstract deictic gesture (i.e. by pointing and thus assigning a space for the referent in the abstract). In the case of Turkish speakers, Azar et al. (2019) observed a tendency for gesture frequency to increase with (1) re-introduced rather than with maintained referents, and (2) with full NPs rather than with pronouns. Debreslioska & Gullberg (2019) report less accessible referents that are expressed with NPs rather than pronouns to occur more frequently with a co-speech gesture, and Debreslioska & Gullberg (2020) find *discourse-given*, but *hearer-new* referents to be more frequently accompanied by a co-speech gesture than all *new* referents. Similarly to Azar et al. (2019), Debreslioska & Gullberg (2022) note a tendency among speakers to gesture along with referents that are being reintroduced, rather than with referents that are maintained. However, in contrast to Azar et al.'s (2019) report, the authors observe this to be the case for German speakers regardless of the lexical form with which the referent is expressed (full NP or pronoun).

In a corpus-based dissertation, Rohrer (2022) finds that co-speech gestures are more likely to co-occur with *new* referents than with *given* ones. Additionally, the author notes that *accessible* referents are more likely to co-occur with a co-speech gesture than *given* referents are, when both are realized with a pre-nuclear (a relatively less prominent) pitch accent. Similarly, sharing vs. not sharing a common ground between the speaker and listener has an effect on gesture frequency. Holler and Stevens (2007) report that information already shared between speakers (i.e. is in the common ground) is less likely to be represented in a co-speech gesture

than is new information, and Jacobs & Garnham (2007) and Galati and Brennan (2014) find speakers to gesture less when retelling a story to an old audience (with whom they share more of a common ground) than with a new audience (with whom they share less of a common ground). In contrast to this, Holler et al. (2011) report that while the number of words and co-speech gestures both *decrease* over time as more of a common ground is shared between interlocutors, there is an *increase* in the rate of co-speech gestures relative to speech, and de Ruiter et al. (2012) and Hoetjes et al. (2015) find no significant differences in gesture rate as a function of givenness, except de Ruiter et al. (2012) observe a decrease in the rate of deictic (pointing) gestures.

1.4.2.2.2 Co-speech Gesture articulation

With respect to the articulation of co-speech gestures (rather than their co-occurrence patterns), Gerwing (2003) and Gerwing & Bavelas (2004) find that even when the co-speech gesture is categorically the same, versions of gestures accompanying information already in the common ground (*given* information) are smaller, less precise, and less complex than those accompanying *new* information. *New* information is thus emphasized by larger and more precisely articulated co-speech gestures. In a similar vein, Galati & Brennan (2014) report co-speech gestures to be smaller and less precise when more of a common ground is shared between participants. These findings demonstrate that the information status of the lexical affiliate affects the phonetic realization of co-speech gestures, indicating a link in their co-production. (Extending the concept of *givenness* to co-speech gestures (that is, with the gesture itself becoming *given*), Holler et al. (2022) find that co-speech gesture articulation indeed changes over time: with multiple repetitions, co-speech gestures become significantly shorter.)

From this subsection it may be concluded that pragmatically prominent units often serve as anchors to co-speech gestures, *focus* and to a certain extent *topic* being the main attractors, and the alignment between the speech unit and the co-speech gesture may be modulated by the relative pragmatic and prosodic prominence of the speech unit, which in turn may depend on a composite factor of (1) type of focus, (2) number of pragmatic functions of the focus, and (3) the information status of the expression and the co-speech gesture. We may also note that the relative prominence of co-speech gestures - measured both phonetically and perceptually - can co-vary with the relative pragmatic and prosodic prominence of their lexical affiliate. This concludes **section 1.4** on co-speech gestures and information structure. **Section 1.5** combines insight from the prosody, information structure, and co-speech gesture literature to introduce the emerging research questions.

1.5 Research Questions

1.5.1 Speech and co-speech gesture coordination

Prosodically prominent units are aligned with co-speech gestures, and the manner in which they align seem to depend in part on the prosodic correlates of prominence in the given language. Under the hypothesis that speech and co-speech gesture alignment is driven by prosodic structure (see Prieto, Esteve-Gibert, and Shattuck-Hufnagel, 2025), inclusion of typologically distinct languages is not only for descriptive purposes: if prosodic structure is the driving factor, the lack of prosodic diversity necessarily limits the scope of our understanding. The majority of the studies reviewed elicit prominence via focus and examine languages that mark focus primarily by pitch accentuation. Few studies examined alignment in languages that mark focus by phrasing and/or by word-order modulation. The question arises therefore, of how

co-speech gestures align with speech in such languages: what exactly are these co-speech gestures coordinated with?

Further, focus often serves as the landing site of co-speech gestures, with types of focus potentially having an effect on gesturing patterns (as found in Carignan et al., 2024). However, focus is not the sole attractor of co-speech gestures, as topic (Türk, 2020) and background units in general (Kügler & Gregori, 2023) have been found to be accompanied by co-speech gestures as well. Despite their proposed lack of active informational function (McNeill, 1992), background units can be prominent enough to anchor gestures. A question arising from these observations, therefore, is how co-speech gestures align with different types of pragmatically prominent units, and what role prosody plays in this alignment.

The question of how speech and co-speech gestures align in a language where the primary expression of information structure is word order modulation can be approached with Hungarian. Focus and topic in this language are expressed by movement into canonical positions in the left periphery (different positions for focus and topic are assumed). In addition to this, focus prominence can be expressed via additional pitch accentuation (Genzel et al., 2015; Mády, 2015; Mycock, 2010) and phrasing, i.e. by the insertion of prosodic boundaries (Mády et al., 2016; Mády & Kleber, 2010). Given the systematic expression of focus via word order and its variable prosodic expression in Hungarian, the question of the nature of alignment emerges for this language: Within a speech unit, what are co-speech gestures most stably coordinated with? Further, does coordination change if the prosodic marking changes? Does Hungarian pattern with Germanic languages in the form of gestural peak to pitch accent peak alignment, or does the co-speech gesture align with some other speech landmark?

Regarding topic prominence, Genzel et al. (2015) consider topics to be part of the background (being non-focus and often *given*), yet topics can be accompanied by prosodic prominence. Given the observation that topics in particular (Türk, 2020), and background units in general (Kügler & Gregori, 2023) can be landing sites for co-speech gestures, it is not improbable that Hungarian topics can be accompanied by co-speech gestures as well. A question arising from Kügler & Gregori's (2023) report of a less tight alignment between pitch accents in the background relative to alignment under focus is whether a similarly less tight alignment can be observed in Hungarian topics. Of course, a "tighter or looser" coordination relative to focus is only possible if systematicity is found to begin with.

A further question is prompted by the previously observed differences in focus and topic prominence. In Hungarian, foci generally have a high pitch accent followed by a fall (Genzel et al., 2015; Mády, 2015), whereas topics may have a rising contour, especially when given (Genzel et al., 2015) and also often when contrastive, with the end of the rise following the stressed syllable given a polysyllabic word (Gyuris & Mády, 2014). What might co-speech gestures coordinate with given such foci and topics? Do we find co-speech gesture peak to tonal peak alignment, as in English, German, French, and Catalan? And with regards to tonal peaks, are co-speech gestures aligned with f_0 maxima regardless of accent type, or with a pitch accent target? In foci, a likely prosodic anchor may be a H tone on the stressed syllable, and in topics with a rising contour, the stressed first syllable or the end of the rise on a subsequent syllable (given a polysyllabic topic) may align with a gesture apex.

The objectives are to explore how speech and co-speech gestures coordinate in Hungarian and to identify which co-speech gestural landmarks coordinate with which phrasal and/or prosodic landmarks. In addressing the question of coordination, the dissertation follows the

concept of *gestural coordination*, in which a regular temporal lag between two gestural events indicates systematicity in their production, called a coupling relationship (see also Leonard & Cummins, 2011).

Although originally used for defining the coordination of C and V gestures in syllables, the concept has been applied at the post-lexical level previously to capture the coordination patterns on intonational landmarks with speech gestural landmarks (see e.g. Mücke et al., 2012), and it has been suggested to include co-speech gestures as well (Krivokapić et al., 2017, Leonard & Cummins, 2011). The findings in Mücke et al. (2012) support the language (and dialect) dependent nature of coordination at the post-lexical level. Intonational tone gestures (H, L) are not part of the lexical representation of words in non-tone languages, like Hungarian, but they do enter into a kind of post-lexical coordination with articulatory gestures. Similar questions may be asked in relation to co-speech gestures (as was done in Krivokapić et al. (2017) for English), which are also not part of the lexical representation of words. Given a tight relationship with prosodic structure, co-speech gestures may coordinate with f_0 (**Figure 2/a**). Given the primary syntactic marking of information structural prominence in Hungarian, the co-speech gesture may coordinate not with f_0 but with some alternative, stable, prosodic landmark of the language. One candidate is the nucleus of the stressed syllable (**Figure 2/b**), which might indicate coordination with a μ -gesture (see Rochet-Capellan et al., 2008; Fuchs & Reichel, 2016 for coordination with lexical stress). As lexical stress falls on the first syllable in Hungarian, coordination with its nucleus might indicate coordination with a prosodic boundary.

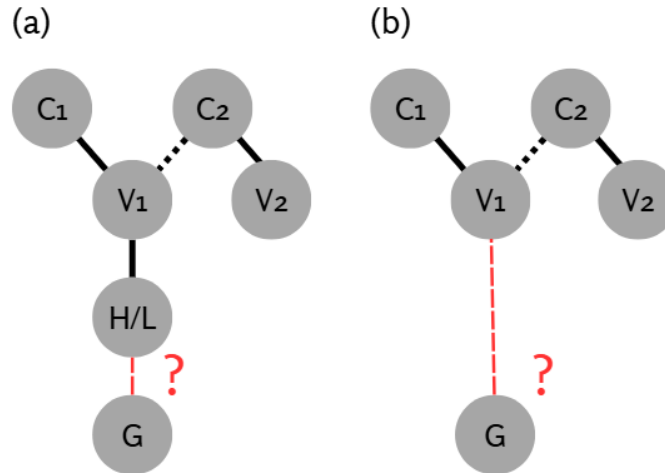


Figure 2 Coupling relationships in a CVCV word between articulatory (C, V), tone (H/L), and co-speech (G) gestures. Solid lines indicate in-phase coupling, dotted lines indicate anti-phase coupling (see Goldstein et al., 2009; Nam, Goldstein, and Saltzman, 2009; Mücke et al., 2012). Red lines with a question mark indicate an unknown coupling relationship.

In summary, the dissertation asks what the linguistic principles behind coordination are.

The specific research questions are listed below.

[RQ1] Which prosodic or articulatory landmarks anchor co-speech gestures to speech? In other words, what are the coordinating landmarks in the alignment of speech and co-speech gestures in Hungarian?

Hypothesis: If speech and co-speech gestures are coordinated with each other, stable coordination will be observed between a speech and a co-speech gestural landmark. Based on prior research, a prosodic landmark can be expected to anchor the gestural landmark.

[RQ2] Do information structural differences modulate the coordination of speech and co-speech gestural landmarks in Hungarian?

Hypothesis: If coordination is driven by prosodic structure and is more stable under greater prominence in languages that mark prominence by pitch accentuation, we will observe more stable coordination between the co-speech gestural landmark and

the coordinating landmark of the pragmatically more prominent speech unit that the gesture accompanies in Hungarian as well. That is, coordination will be more stable under pragmatically more prominent lexical items relative to less prominent ones: (1) focus vs. background, (2) corrective vs. confirmation vs. broad focus, (3) syntactically marked narrow focus vs. syntactically unmarked broad focus.

An additional exploratory question may be addressed if the data is such that statistical analysis can be performed. Boundaries may have an indirect effect on coordination. In Hungarian, a high tone may occur on the word *preceding* the first pitch accent of the sentence (see Varga, 2002), which is expected on the topic or focus. The phonological status of this tone is uncertain as there is no ToBI system in place for Hungarian yet, but in Varga (2002) the term *preparatory* tone is adopted. The question that arises is whether this initial preparatory tone, when present, has a similar effect on tone alignment as found in Esteve-Gibert & Prieto (2013) for boundary tones; namely, in the case of Hungarian, a *rightward* movement of the target-word pitch accent that follows the pretarget-word preparatory tone to avoid tonal crowding. If the co-speech gesture target coordinates with the pitch accent target, it ought to also move in a rightward direction – preserving the stable lag between the pitch accent and co-speech gesture targets.

This effect may be observed when a *word boundary* precedes the lexical affiliate, as in other cases, there is a high likelihood of a longer pause between the boundary tone of the preceding word and the pitch accent on the lexical affiliate, resulting in little chance of tonal crowding.

Further, given the potential difference in pitch accentuation on topic and focus, an indirect IS effect may also be observed, such that we may see a rightward movement in focus conditions (the target of which may likely be reached within the stressed syllable of the lexical

affiliate) and no such effect in topics (the target of which may likely be reached at the end of the lexical affiliate). **Figure 3** below illustrates the possible effects of a preparatory tone given an early peak in focus and a late peak in topic, contrasting also word boundary and IP boundary conditions.

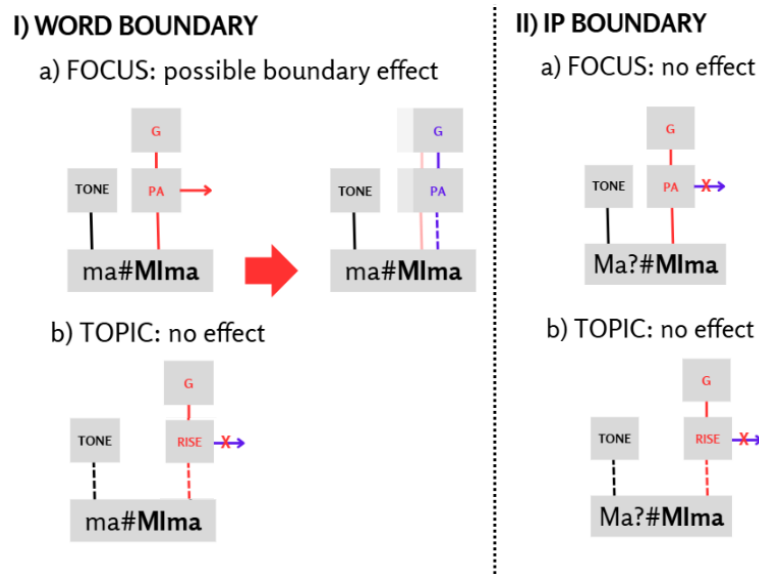


Figure 3 Potential effect of a preparatory tone on tonal alignment given a word boundary preceding the target (vs. no effect following a bigger boundary).

1.5.2 Prosodic boundary and prominence effects

As mentioned in the literature review, there is some evidence that the temporal properties of co-speech gestures change as a function of prosodic structure and prominence. In particular, Krivokapić et al., (2017) finds that co-speech gestures, similarly to speech gestures, become increasingly longer under prosodic boundaries of increasing levels, or in AP terms, under increased activation of the π -gesture (Byrd & Saltzman 2003). Gestures co-active with the π -gesture lengthen; more specifically, the *parts* of a gesture co-active (overlapping) with the π -gesture lengthen. The amount of lengthening is directly proportional to the level of activation of

the π -gesture during the overlap and the greater the overlap between the π -gesture and the gesture, the greater the lengthening effect (see **Figure 4**).

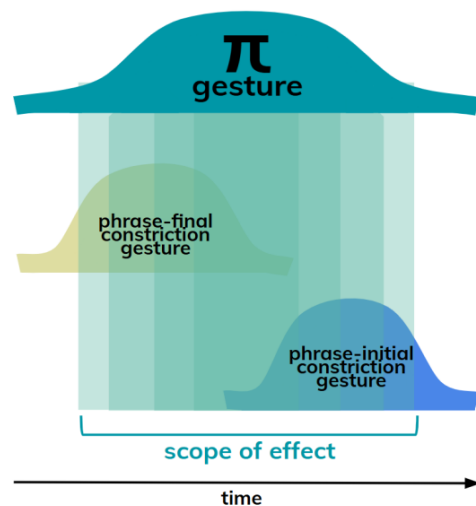


Figure 4 Schematic representation of the scope of the π -gesture's lengthening effect on co-active gestures. Darker to lighter shading indicates the decrease in effect. Figure adapted from Byrd & Krivokapić (2021)

Boundary-related lengthening of speech segments has been reported for Hungarian (e.g. Gósy & Krepsz, 2018; White & Mády, 2008), and the primary question related to prosodic boundaries is whether we observe related lengthening of co-speech gestures as well in Hungarian. If so, that would provide evidence that linguistic structure affects co-speech gestures; i.e. that they are tightly connected to prosody not only by coordinating with prominent syllables, but also by their integration with prosodic structure.

In addition to boundary-related lengthening, Krivokapić et al. (2017), Rusiewicz et al (2014), and Parrell et al (2014) find prominence-related lengthening of co-speech gestures. Lengthening due to prominence is modeled in Articulatory Phonology as the effect of a temporal modulation gesture called the μ -gesture (Saltzman et al., 2008). The effect of prominence is expected to be the strongest at the syllable nucleus and taper out away from it (Bombien et al.,

2013). There is some evidence for prominence having a modulating effect on vowel duration in prominent syllables in Hungarian as well (Markó et al., 2018), although Vogel, Athanasopoulou, & Pincus (2015) and White & Mády (2008) report evidence to the contrary. Therefore, if a μ -gesture is indeed present on the prominent syllable (as would be evidenced by the lengthening of the vowel), and if co-speech gestures are affected by the μ -gesture, lengthening of co-speech gestures due to prominence may then be expected on phases of the gesture that overlap with the nucleus of the prominent syllable. In addition to the possibility of lengthening, the insertion of prosodic boundaries has been noted as a possible marker of prominence in Hungarian (Mády, Kleber, Reichel, & Szalontai, 2016). Should this be observed, the effect of those boundaries on the temporal properties of co-speech gestures can be measured as well.

To summarize, the dissertation examines the following questions related to prosodic boundary and prominence effects of the temporal properties of pointing gestures:

[RQ3] How do prosodic boundaries affect co-speech gestures?

Hypothesis: If co-speech gestures are affected by prosodic boundaries (i.e., the π -gesture), a lengthening effect on parts of the co-speech gesture that overlap with a prosodic boundary will be observable.

[RQ4] How does prominence affect the temporal properties of co-speech gestures?

Hypothesis: If prominence affects the temporal properties of speech and co-speech gestures, information structure corresponding to greater pragmatic prominence will have a lengthening effect (corrective focus, contrastive topic > broad focus; corrective focus > contrastive topic).

[RQ5] How does the interaction of prosodic boundaries and prominence influence the temporal properties of co-speech gestures?

Hypothesis: *If prosodic boundaries and prominence both have a lengthening effect on co-speech gestures, a cumulative lengthening effect will be observable with more lengthening under a prosodic boundary given greater prominence than given smaller prominence.*

1.5.3 Proposed study and dissertation chapters

The dissertation project addresses the research questions laid out above via a semantically-pragmatically controlled experiment utilizing Electromagnetic Articulography (EMA). EMA is a tool that measures the position and movement of the articulators (such as the tongue and the lower lip) in speech production via sensors. Given its high temporal resolution, EMA is exceptionally well-suited for analyzing the precise coordination and temporal properties of speech and co-speech gestures, thus providing an ideal methodology for the present project. In addition to articulatory data, concurrent visual and acoustic data were collected as well.

The dissertation chapters are organized as follows: ***Chapter 2*** details the methodology of the dissertation; specifically, it describes the study design and data collection procedure that apply to both the speech and co-speech gesture coordination and lengthening questions. ***Chapter 3*** presents the specific methodology, results, and discussion related to the question of coordination (RQ1-2). ***Chapter 4*** describes the methodology, results, and discussion related to the lengthening question (RQ3-5). ***Chapter 5*** consists of an overview and discussion of the overall pattern of results and their implications to the interrelationship of prosody, information structure, and co-speech gestures.

Chapter 2 Methodology

The present chapter describes the methodology of the dissertation, addressing the two sets of research questions: (RQ1-2) speech and co-speech gestures coordination and (RQ3-5) the temporal properties of co-speech gestures under prosodic boundaries and pragmatic prominence in Hungarian. The chapter focuses on those aspects of the methodology that are shared by the two studies (participants, recording procedure, experimental tasks, data processing and labeling procedure), and the following two chapters detail additional aspects of the studies pertinent to answering the two RQs (specific sets of stimuli materials, data analyses, statistical analyses).

2.1 Participants

Seven adult native speakers of Hungarian (4m, 3f) participated in the study. All participants grew up in Hungary in majority Hungarian-speaking communities with varied exposure to other languages taught in school.

Participants' region of origin was not restricted to any particular area of Hungary. Following Schiffman's (2017) definition, Hungary can be characterized as a country with *diglossia*: “[d]iglossic languages are usually described as consisting of two (or more) varieties that coexist in a speech community; the domains of linguistic behavior are parceled out in a kind of complementary distribution” (Schiffman, 2017:205). Most speakers of Hungarian speak at least two varieties of Hungarian: a standard (also called *művelt köznyelv*, “Educated Colloquial Hungarian”) and a regional dialect (Nagy & Császi, 2015). The prevailing distinction in the spaces of use between the standard and a dialect is on the spectrum of *formality* to *familiarity* of

the given situation (Kiss, 2006), with the standard being used in formal situations (e.g. school setting) and a dialect in familiar ones (e.g. home setting). For this reason, given the formal environment of the study, speakers used Educated Colloquial Hungarian throughout our interactions as well as during recording.

At the time of data collection, participants had lived in the United States for varied durations (three participants for under a year, one participant for six years, one participant for eight years, and two participants for nine years). Participants reported no speech impediments or hearing loss. They were compensated for their time and effort.

2.2 Materials

2.2.1 Experimental conditions

The EMA study examines the effect of pragmatic and prosodic prominence on gestural coordination, as well as boundary effects on speech and co-speech gestures. **Table 1** below illustrates the conditions.

For the question of the effect of information structure on coordination, five pragmatic prominence conditions were included: broad focus (#4), aboutness topic (#5), contrastive topic (#6), corrective focus (#7), confirmation focus (#8). To examine the effect of prosodic boundaries, a baseline broad focus condition and a corrective focus condition were included at two boundary levels, word (#4, #7) and Intonational Phrase or IP (#9, #11). For the potential effect of a preparatory tone on tonal and gestural alignment (see Esteve-Gibert & Prieto, 2013), topic and focus conditions were collected in word and IP boundary conditions: contrastive topic in word condition (#6), contrastive topic in IP condition (#10), corrective focus in word condition (#7), corrective focus in IP condition (#11).

Table 1 Experimental conditions. Dark gray shading indicates that no data was collected for the cell. Light gray shading indicates conditions not examined in the dissertation.

INFORMATION STRUCTURE CONDITIONS		GESTURE CONDITIONS		
		NO GESTURE	GESTURE	
			PROSODIC BOUNDARY CONDITIONS	
			WORD	IP
BASE	FOCUS: SECOND- OCCURRENCE BROAD	#1		
	FOCUS: BROAD	#2	#4	#9
TOPIC	TOPIC: ABOUTNESS		#5	
	TOPIC: CONTRASTIVE		#6	#10
NARROW FOCUS	FOCUS: CORRECTIVE	#3	#7	#11
	FOCUS: CONFIRMATION		#8	
	FOCUS: CONTRASTIVE*		#12	
	FOCUS: QNA*		#13	

Further conditions were included which are reserved for future work and are not examined in this dissertation. The first set of conditions correspond to the questions regarding the role of information structure; these are non-corrective narrow (“QNA”, #13) and contrastive focus (#12). To examine the effect of co-speech gestures on the coordination of tone and constriction gestures in the future, *gesture vs. no gesture* conditions were included: second-occurrence broad focus without gesture (#1) to examine the coordination of constriction gestures without expected intonational movement or pitch accentuation, broad focus without gesture (#2) to examine the coordination of an F0 target with constriction gestures, and broad focus with gesture (#4), to examine the effect of co-speech gestures on the coordination of an F0 target with constriction gestures. (As a backup, corrective focus with no gesture (#3) was included as well for this latter question.)

2.2.1.1 Stimuli: Information structure

A target utterance, shown below in (1) and (2) was constructed to be used in every condition. Note that in Hungarian, as described in **Chapter 1**, the sentence necessarily has an alternative form: although in general, the verbal prefix (*vpx*) precedes the verb (as in (1)), in narrow focus, it appears post-verbally (as shown in (2)). The information structure conditions in the study that are affected by this are confirmation and corrective focus. Importantly, the verbal prefix-verb combination was chosen so that in both syntactic structures, the target word (*Mima*) is followed by the syllable *be-*.

- [1] *Ma Mima be- betonozta a hírnevét.*
Today Mima.NOM vpx-cemented.3pSg the reputation.3pPoss.ACC
“Today, Mima cemented her reputation.”

- [2] *Ma Mima betonozta be a hírnevét.*
today Mima.NOM cemented.3pSg vpx the reputation.3pPoss.ACC
“Today, Mima cemented her reputation.”

The target sentence followed a discourse context (a background story of a few sentences), and it functioned as the answer to context-setting questions that corresponded to the discourse context. The target sentence was constructed with a contextually given target word (*Mima* ['mimɒ]). The question-answer pairs were designed to elicit the correct information structure (topic and focus types) on this target word. The discourse context and the question-answer pairs are given in **Table 2** below. Note that in the contrastive topic condition, the target sentence was followed by another sentence corresponding to the second topic phrase being contrasted with

Mima (Question: *What happened at the competition with **you and Mima** today?* Answer: *Today/Today? **Mima** cemented her reputation. And **I** just admired all the skillful work.)*

The discourse referents that appear in the questions and answers are presented first in the discourse context. Importantly, the background story introduces the name (and target word) *Mima*, which does not exist as such in Hungarian. Two other names used in the background story and in the context-setting questions, *Papi* and *Barbi*, exist in the language: *Papi* is a commonly used affectionate term for *nagypapa* “grandpa”, and *Barbi* is a nickname for *Barbara*.

The effects of givenness are not directly targeted in this design and all discourse referents are contextually *given* in the sense that they are introduced in the background story. In the topic conditions, *Mima* is additionally set up in the question to serve as a topic in the answer. In the confirmation focus condition, *Mima* is *given-as-focused*, with the understanding that the question introduces *Mima* as being in focus, whereas in the corrective focus condition, *Papi* is given in the question as being in focus, and *Mami* serves as a corrective replacement. In the broad and corrective focus conditions, *Mima* is not additionally mentioned in the question.

Table 2 Stimuli conditions: green indicates conditions analyzed in the dissertation, yellow indicates data collected not as part of the dissertation, and gray shading indicates no data was collected for the cell.

Background story	<i>The story and the corresponding conversations feature Papi, Mima (Papi’s wife) and Barbi (their granddaughter). By the way, Mima’s real name is Emőke, but when Barbi was born, everyone in the family started calling her Emi mama; however. Barbi could not pronounce that as a toddler, so she called her Mima instead. The nickname stuck, and everyone calls Emi mama Mima now.</i> <i>Anyhow, what’s important is that the three of them are really into embroidery. Mima has been a member of an embroidery club for years, Papi joined fairly recently as well, but Barbi is too young to join. This week is the annual two-day competition of the club, where a winner is announced each day. Papi and Barbi discuss the events of the competition.</i>
------------------	---

Target sentence	Ma Mima (be-) betonozta (-be) today Mima.NOM VPX-cemented.3pSg -VPX a hírnevét. the reputation.3pPOSS.ACC “Today, Mima cemented her reputation.”	Ma? Mima (be-) betonozta (-be) today Mima.NOM VPX-cemented.3pSg -VPX a hírnevét. the reputation.3pPOSS.ACC “Today? Mima cemented her reputation”
Prosodic boundary	WORD	IP
base – second-occurrence broad focus	B: What happened at the competition today? P: <i>Ma Mima bebetonozta a hírnevét.</i> “Today Mima cemented her reputation.” B: <i>Tudom, hogy ma Mima bebetonozta a hírnevét, de ki nyerte meg a versenyt?</i> “I know that Mima cemented her reputation today, but who won the competition?”	
base - broad focus	What happened at the competition today? P: <i>Ma Mima bebetonozta a hírnevét.</i>	What happened at the competition today? P: <i>Ma? Mima bebetonozta a hírnevét.</i>
topic - aboutness	B: What happened at the competition with Mima? P: <i>Ma Mima bebetonozta a hírnevét.</i>	
topic - contrastive	B: What happened at the competition with you and Mima? P: <i>Ma Mima bebetonozta a hírnevét.</i>	B: What happened at the competition with you and Mima today? P: <i>Ma? Mima bebetonozta a hírnevét. Én pedig csak csodáltam a sok ügyes munkát.</i> [Today? Mima cemented her reputation. And I just admired all the skillful work]
focus - agreement	B: Was Mima exceptionally good today? P: <i>Ma Mima betonozta be a hírnevét.</i>	

focus - corrective	B: Was Papi exceptionally good today? P: <i>Ma Mima betonozta be a hírnevét.</i>	B: Was Papi exceptionally good today? P: <i>Ma? Mima betonozta be a hírnevét.</i>
focus - Q&A	B: Who stood out today? P: <i>Ma Mima betonozta be a hírnevét.</i>	
focus - contrastive	B: Was either of you (you or Mima) exceptionally good today? P: <i>Ma Mima betonozta be a hírnevét.</i>	

As mentioned above, the target word in each sentence is “Mima”. The consonant /m/ was chosen for ease of tracking in EMA, and the vowels /i/ and /ɒ/ were chosen for the same reason, as they are produced in opposite areas in the vowel space (high front and low back, respectively), and the tongue is expected to move from one to the other in a trackable manner. The segments preceding and following the target word alternate in their vowels and consonants in a similar fashion: *ma Mima be* [ˈmɒ ˈmimɒ ˈbɛ]. The verb *betonozta* (cemented.3pSg) was chosen to match in its first syllable with a contextually appropriate prefix (*be-*), so that in both the topic and broad focus, as well as the narrow focus conditions, the target word is followed by the same consonant-vowel sequence. The whole sequence *ma Mima be* is voiced to allow for the labeling of f0. Lexical stress falls on the first syllable of the target word.

2.2.1.2 Stimuli: Prosodic boundary conditions

As shown in **Table 2**, to elicit the appropriate prosodic boundaries, the target word is preceded by (a) a word boundary in *Ma Mima bebetonozta a hírnevét* (topic and broad focus on *Mima*) and *Ma Mima betonozta be a hírnevét* (narrow focus on *Mima*) “Today, Mima cemented her reputation”, and (b) an IP boundary in *Ma? Mima bebetonozta a hírnevét* (topic and broad focus on *Mima*) and *Ma? Mima betonozta be a hírnevét* (narrow focus on *Mima*) “Today? Mima

cemented her reputation.” The target word in each condition is controlled to be followed by a word boundary.

2.2.2 Tasks

Participants were asked to first read the situating story silently. The story was followed by question-answer pairs. Participants were told that each question corresponds to the same overarching story, but the different question-answer pairs are not related to each other. They were asked to read the pairs first silently and then aloud. Images of the characters mentioned in the story were displayed in front of the participants. In the gesturing conditions, participants were asked to point at the picture corresponding to the character mentioned in the target sentence. In the read-only conditions, participants were not instructed to point. The cameras were placed at a distance that adequately captured the resting position as well as the full range of pointing motion of the participant. An external microphone was placed in front of the participant.



Figure 5 Experimental setting

Each target sentence was produced 10 times by each participant, resulting in 110 utterances per participant from the five pragmatic, two boundary, and gesture-no gesture conditions. 20 further utterances were produced due to the additional focus conditions that are not analyzed as part of the dissertation. Filler sentences with a different target word were added to diffuse attention from the task of pointing at one target word. To reduce the cognitive effort required in reading and interpreting the questions for the correct information structural productions, the question-answer pairs were blocked into two pragmatic conditions at a time, with the tokens presented in a randomized order within each block. To reduce instruction time, conditions in which participants were not asked to gesture (“READ”) were blocked together separately from conditions in which they were asked to gesture (“READ-GESTURE”). Blocks contained either 6 or 4 repetitions of tokens per pragmatic condition, and each block with 6x2 tokens had a sister block with 4x2 tokens, resulting in 10 tokens per pragmatic conditions total (see **Table 3** for details). To break up the monotony of the task, participants were asked to do additional tasks between certain blocks (such as pointing at filler pictures in the room).

Table 3. Experimental blocks

Information Structure	Prosodic Boundary	Repetition	Block
Instructions given & background story presented			
Focus: SO broad	Word	6x	READ 1
Focus: corrective	Word	6x	
Focus: broad	Word	6x	READ 2
Filler sentence			
Instructions given & background story presented			
Focus: broad	Word	6x	READ-GESTURE 1
Focus: corrective	IP	6x	
Focus: broad	IP	6x	READ-GESTURE 2
Focus: confirmation	Word	6x	
Filler pointing task			

Information Structure	Prosodic Boundary	Repetition	Block
Instructions given & background story presented			
Focus: corrective	Word	6x	READ-GESTURE 3
Focus: contrastive	Word	6x	
Focus: QNA	Word	6x	READ-GESTURE 4
Topic: contrastive	IP	6x	
Focus: contrastive	Word	6x	READ-GESTURE 5
Topic: aboutness	Word	6x	
Filler sentences			DUMMY BLOCK
Filler pointing task			
Instructions given & background story presented			
Focus: broad SO	Word	4x	READ 3
Focus: corrective	Word	4x	
Focus: broad	Word	4x	READ 4
Filler sentence			
Instructions given & background story presented			
Focus: broad	Word	4x	READ-GESTURE 6
Focus: corrective	IP	4x	
Focus: broad	IP	4x	READ-GESTURE 7
Focus: confirmation	Word	4x	
Filler pointing task			
Instructions given & background story presented			
Focus: corrective	Word	4x	READ-GESTURE 8
Focus: contrastive	Word	4x	
Focus: QNA	Word	4x	READ-GESTURE 9
Topic: contrastive	IP	4x	
Focus: contrastive	Word	4x	READ-GESTURE 10
Topic: aboutness	Word	4x	
Filler sentences			DUMMY BLOCK

2.2.3 Data collection procedure

Kinematic, visual, and acoustic data were collected at the EMA lab of University of Michigan. Kinematic data was collected with a Carstens Medizinelektronik AG501 3D electromagnetic articulography (EMA) system. EMA sensors were placed on the tongue body

and dorsum to track movement for vowels and on the upper and lower lips to track bilabial consonants. To provide reference points for the head, two sensors were placed on the left and right mastoid processes and one on the upper right incisor. Two cameras recorded each participant; these were placed to record frontal and profile views. An external microphone was placed in front of the participant.

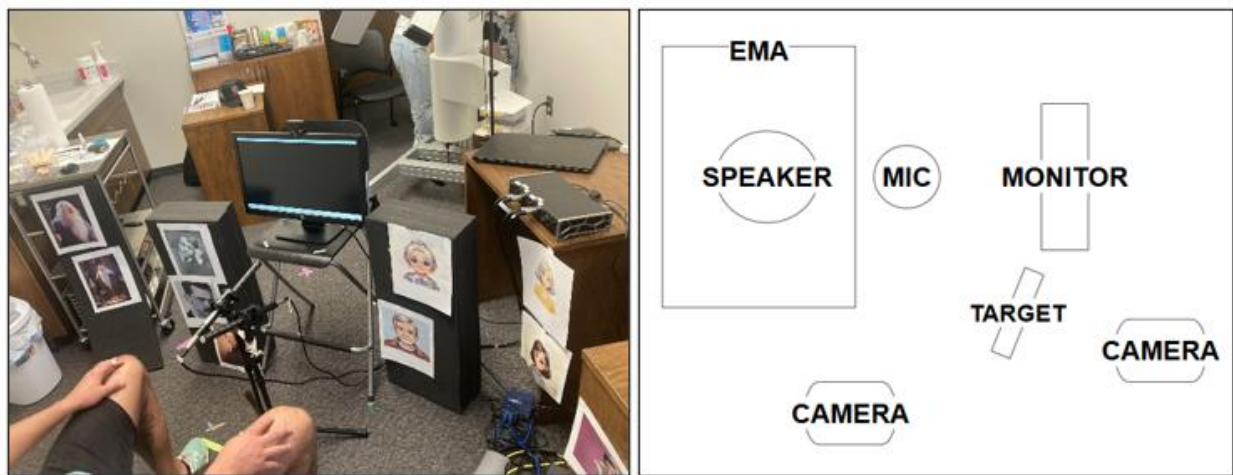


Figure 6. Experimental setup

Prior to recording, the natural resting position of participants' hands were marked with a sticker on their legs. Next, participants were familiarized with the characters in the story by reviewing their names and images, displayed twice on a monitor in front of them. This was followed by a practice session in which participants familiarized themselves with the reading and pointing tasks. During this session, they were asked to read a short context setting story. This was followed by two sets of question-answer pairs related to the story. In the first set, participants were asked to read three pairs. In the second set, they were asked to read the same three pairs and gesture towards the images of the characters mentioned.

The story and the question-answer pairs in the practice session and in the experimental blocks that followed were displayed on the monitor in front of the participant, placed at a

comfortable reading distance, which was confirmed with each participant. Images of the characters were printed out and placed around the participant in the same arrangement for each recording session.

As in the practice sessions, participants read the context-setting story first, then they read the question-answer pairs aloud and pointed at the appropriate picture as they mentioned the character. Participants were asked to rest their gesturing hand on their corresponding leg where the sticker was placed as a reference point when not gesturing (see Rochet-Capellan et al., 2008; Esteve-Gibert & Prieto, 2013; Krivokapić et al., 2017). The two cameras were placed at a distance that adequately captured the resting position as well as the full range of pointing motion of the participant.

2.2.4 Data processing and annotation

Utterances with disfluencies and speech errors were excluded, as well as utterances in the READ-GESTURE conditions during which speakers did not produce a gesture. P2's data was excluded due to significant loss of data. The utterances were examined in Praat (Boersma & Weenink, 2024) to ascertain that the sentences were produced with the expected prominence and boundary type.

2.2.4.1 Constriction and pointing gestures

The video recordings of P1-P6 were time-aligned with the acoustic signal and kinematic data using *Adobe Premiere Pro 2025*. Due to software error, P7's data was aligned using a custom *MATLAB* script by Yue Liang (University of Michigan). After alignment, co-speech gestural data was extracted from the visual data using the full body tracking module of *EnvisionBox* ("Media-Pipe (and Optionally Masking)", Owoyele et al., 2022). The module inputs

video files and produces kinematic time-series data as output, tracking the movement of points on the hands, face, and body in three dimensions (x, y, z axes per frame over time). As participants were only instructed to point toward the target, but were not told how to gesture, the gestures produced were variable; therefore, the relevant tracking dimensions (vertical or horizontal) were determined for each participant separately. The point corresponding to the tip of the gesturing index finger was the most informative articulator in each case (see **Table 4**). The co-speech gesture data stream was combined with the kinematic data from EMA using a custom script by Cynthia Lee (University of Southern California).

Table 4. Pointing gesture dimensions labeled

Participant	Gesturing hand	Dimension
P1	right	horizontal (forward pointing movement)
P2	<i>Excluded due to data loss.</i>	
P3	right	vertical (downward pointing movement)
P4	left	horizontal (forward pointing movement)
P5	predominantly right	vertical (downward pointing movement)
P6	right	vertical (downward pointing movement)
P7	right	vertical (downward pointing movement)

The kinematic data were annotated using *mview* (software created by Mark Tiede, Haskins Laboratories/Yale University), see **Figure 7**. Consonants and vowels were annotated using the *FindGest* labeling method. Consonants were labeled on the lip aperture trajectory (LA, the Euclidean distance between upper and lower lip markers) in the target word and in the pre- and post-target bilabial consonants. Vowels were labeled on the vertical movement of the tongue dorsum in the target word (TD), as the dimension was salient and reliably labelable. Timepoints

extracted for speech gestures were gesture onset (*GONS*), peak velocity (*PVEL*), nucleus onset (*NONS*), maximum displacement (*MAXC*), nucleus offset (*NOFFS*), peak velocity (*PVEL2*) and gesture offset (*GOFFS*).

For co-speech gestures the vertical or horizontal movement of the pointing index finger was labeled in *mview*; the dimension labeled depended on the salient trajectory of the finger movement (participants either pointed in an up-down motion or in a forward motion). Two different labeling methods were necessary due to the shape of the pointing gesture trajectories. For speakers P1, P3, P6, and P7, *FindGest* was used, producing the same landmarks as in the case of the constriction gestures. Two speakers, P4 and P5, produced larger preparatory and return-to-rest movements before pointing, making the labeling of the smaller pointing gesture in-between impossible using *FindGest*. For these two speakers, the *snapex* labeling method was used instead to label the local minimum and maximum points in a selected trajectory (the local minimum is the onset of the pointing movement, and the local maximum is the target). This alternative method allowed for the labeling of small pointing movements between the larger preparatory and return to rest movements.

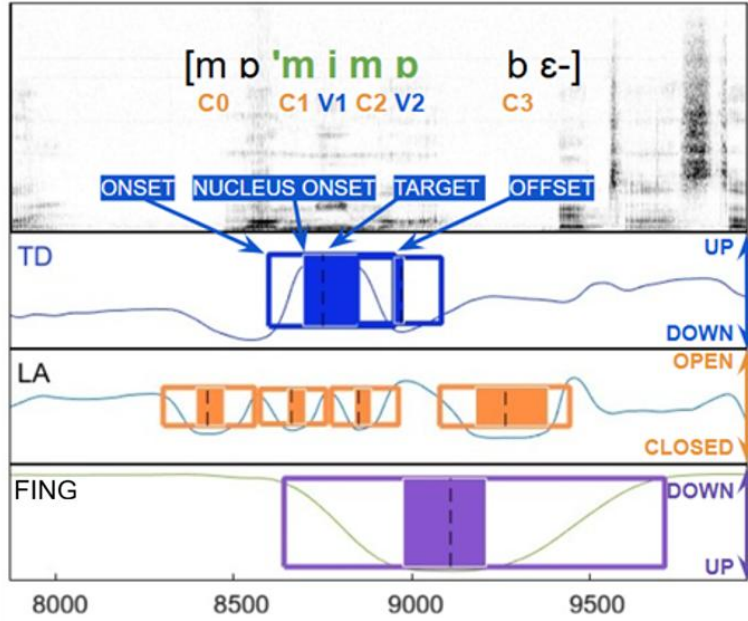


Figure 7. Speech and co-speech gesture labeling in mview: spectrogram, tongue dorsum (TD, vertical) lip aperture (LA), right index fingertip (FING, vertical). The target word ['mimɒ] is highlighted in green. The left end of the box (label) is the gesture onset; the left end of the shaded box is the gesture nucleus onset (onset of plateau); the dotted line is the gesture target; the right edge of the shaded box is the gesture nucleus offset; the right edge of the box is the gesture offset.

2.2.4.2 Prosody

Time-aligned acoustic data was annotated in Praat (Boersma & Weenink, 2024) for intensity, pitch accent onsets and targets, and f0 rise onsets and targets within the target word. Amplitude maxima timepoints associated with the target word were extracted to examine intensity peak and co-speech gesture target coordination (Pouw et al., 2020a). **Figure 8** illustrates pitch labeling. The onset and target of pitch accents associated with the target word were labeled with the following labeling conventions (**Figure 8**, yellow dots). In case of high pitch accents, the target was the maximum f0 of the pitch accent (labeled manually, via the ‘move cursor to maximum pitch’ function) and the onset was labeled manually as the onset of the upward curve (**Figure 8/a**). In cases of low accents (where f0 was flat throughout the first syllable of the target word, where lexical stress and phrasal prominence are expected to be

realized), the target word onset was marked as the onset and the midpoint of the lexically stressed vowel was marked as the peak, being the nucleus of the prominent syllable (**Figure 8/b**). In addition to pitch accents, f0 maxima were also labeled in the target word (**Figure 8**, pink dots). These labels may or may not coincide with the separate pitch accent labels: **Figure 8/a** - in case a high pitch accent was present on the first syllable, the pitch accent and f0 labels coincided; **Figure 8/b** - if the first syllable had flat f0 but the second (=non-prominent) syllable exhibited f0 movement (a rise in all cases), the beginning and end points of the rising contour were marked as f0 onset and peak, respectively -- these labels do not coincide with the low pitch accent label on the first syllable.

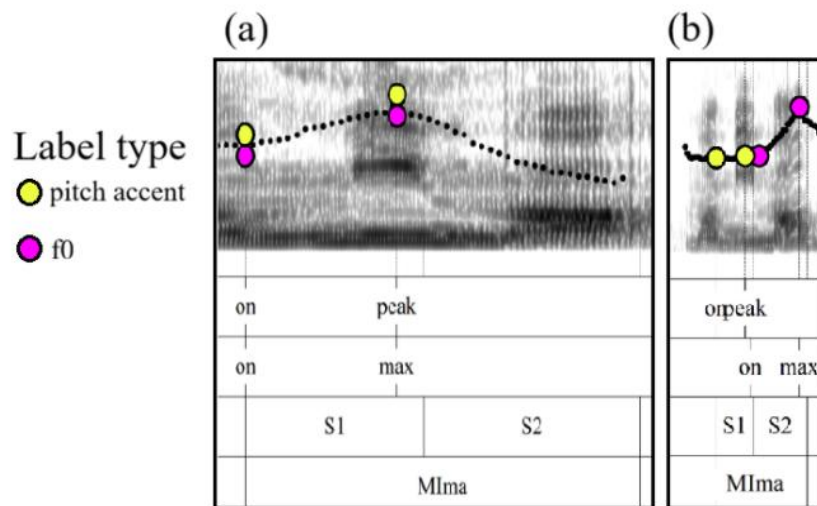


Figure 8. Pitch accent and f0 labeling convention. Panel (a) shows a token with a H pitch accent, where the PA and f0 labels coincide. Panel (b) shows a token with a L pitch accent, where the PA and f0 labels diverge.

To further examine the role of f0 in coordination, the produced contours on the target word were clustered using the contour clustering method of Kaland (2023, 2026), with default settings. To account for speaker variation, the built-in setting for z-score standardization of f0 was used.

Chapter 3 Coordination

3.1 Introduction and research questions

A central question in the relationship between speech and co-speech gestures is understanding in what way the two are coordinated in speech production. The introductory chapter (*Chapter 1*) surveyed the evidence for the close temporal alignment of prosodic prominence (lexical and post-lexical) and gestural landmarks, as well as for the broad synchronization of gestural and prosodic phrases of varying levels. The chapter also illustrated that the question of alignment has been addressed in numerous studies, thus far in a select number of (mainly typologically close) languages. Many studies were conducted on languages in which prominence is marked via pitch accentuation, and prominence was elicited via focus. Results from this body of research indicate that alignment is driven by prosodic structure. Due to the lack of linguistic diversity, however, our view of how co-speech gestures align or coordinate with speech gestures is necessarily limited. In order to contribute to a better understanding of coordination, more research on typologically distinct languages is needed. The present chapter provides one such study with Hungarian as the object language – a language in which pragmatic prominence is primarily expressed via word-order modulation and can additionally be marked by pitch accentuation and phrasing (Genzel, Ishihara, & Surányi, 2015; Mády, 2015; Langer & Kügler, 2023; Mády, Kleber, Reichel, & Szalontai, 2016; Mády & Kleber, 2010). The specific research questions examined are reiterated below.

[RQ1] Which prosodic or articulatory landmarks anchor co-speech gestures to speech? In other words, what are the coordinating landmarks in the alignment of speech and co-speech gestures in Hungarian?

Hypothesis: If speech and co-speech gestures are coordinated with each other, stable coordination will be observed between a speech and a co-speech gestural landmark. Based on prior research, a prosodic landmark can be expected to anchor the gestural landmark.

[RQ2] Do information structural differences modulate the coordination of speech and co-speech gestural landmarks in Hungarian?

Hypothesis: If coordination is driven by prosodic structure and is more stable under greater prominence in languages that mark prominence by pitch accentuation, we will observe more stable coordination between the co-speech gestural landmark and the coordinating landmark of the pragmatically more prominent speech unit that the gesture accompanies in Hungarian as well. That is, coordination will be more stable under pragmatically more prominent lexical items relative to less prominent ones: (1) focus vs. background, (2) corrective vs. confirmation vs. broad focus, (3) syntactically marked narrow focus vs. syntactically unmarked broad focus.

The objectives are thus to explore how speech and co-speech gestures align in Hungarian, focusing on pointing gestures, by identifying which co-speech gestural landmarks coordinate with which speech landmarks, and by examining how that coordination may be modulated by pragmatic prominence corresponding to information structural differences.

3.2 Methodology

The methodology of the study is laid out in *Chapter 2*. Details specific to answering the research question regarding gestural coordination are described below.

3.2.1 Experimental conditions and stimuli

To examine the question of coordination, participants were presented with five types of question-answer pairs, with each question type eliciting different information structure on a specific target word in the answer. In all cases, the target word was *Mima*, which was preceded by the one-syllable word *ma* [“today”] and followed by the verb (*be-*)*betonozta* [(verbal prefix-) “cemented”]. The questions below elicited different information structures on the target word in the answer. English translations of the stimuli relevant to the present chapter are listed below. The full list of stimuli was given in *Chapter 2*.

[a] Questions:

- a. Broad focus: “What happened at the competition today?”
- b. Confirmation focus: “Was it Mima that was exceptionally good today?”
- c. Corrective focus: “Was it Papi that was exceptionally good today?”
- d. Aboutness topic: “What happened with Mima at the competition today?”
- e. Contrastive topic: “What happened with you and Mima at the competition today?”

[b] Answer: *Ma Mima (be-) betonozta (-be) a hírnevét.*

today Mima.NOM VPX-cemented.3pSg -VPX the reputation.3pPOSS.ACC

“Today, Mima cemented her reputation.”

In the contrastive topic condition, the answer portion also included reference to what happened to the character answering the question: “Today, Mima cemented her reputation. And I just admired all the skillful work.”

In the broad focus, aboutness topic, and contrastive topic conditions, the target was followed by the *bebetonozta* (verbal prefix – verb) sequence; in the confirmation focus and corrective focus conditions, the sequence *betonozta be* (verb – verbal prefix) followed. In all cases, therefore, the target phonetic sequence was [mɒ mimɒ bɛ].

3.2.2 Data analysis

3.2.2.1.1 Concept and examined landmarks of coordination

The dissertation adopts the concept of *gestural coordination* from Articulatory Phonology (Browman & Goldstein, 1992, 1995; Goldstein, Byrd, & Saltzman, 2006), as applied to post-lexical prosodic gestures and/or co-speech gestures previously in Katsika et al. (2014); Krivokapić et al. (2017), and Leonard & Cummins (2011). As described in **Chapter 1**, prior research suggests that a co-speech gestural landmark is temporally anchored to a speech landmark. To determine which landmarks are thus coordinated in Hungarian speech and co-speech gesture production, we examine the normalized temporal distances between pairs of co-speech and speech gestures with the understanding that the least variable temporal distance among the pairs signifies the correct pair of coordinating landmarks. That is, coordinating landmarks occur at a relatively more stable temporal distance from one another. A related note on terminology: the term *coordination* will be used for gestural coordination as defined in AP, and the term *temporal alignment* will be used pretheoretically.

The intervals examined are listed in **Table 5**. In Articulatory Phonology, a synchronous coordination pattern between onsets (for C and V gestures) is considered the most stable (Browman & Goldstein, 2000; Goldstein, Byrd, & Saltzman, 2006; Nam, Goldstein, & Saltzman, 2009; Nam & Saltzman, 2003; see Byrd & Krivokapić (2021) for an overview), and for this reason, a number of onset-to-onset intervals are included in the analysis corresponding to the temporal distances between the onset of the co-speech gesture and onsets of constriction (C0, C1, V1) and prosodic (PA, f0) gestures. Next, as described in the introduction, prior literature reports the target of the co-speech gesture to be a likely landmark of coordination, often being anchored to pitch accents. This pair of landmarks is thus included in the analysis, along with other

potential target-to-target intervals between the co-speech gesture target and targets of constriction (C0, C1, V1) and prosodic (PA, f0, intensity) gestures.

Table 5. Pairs of landmarks considered for coordination. (Note that the numbers in the coding following the words ONSET and TARGET are for ordering purposes and are not meaningful.)

Speech gesture	Co-speech gesture	Coding
pre-boundary consonant onset (<u>m</u> a Mima)	pointing gesture onset (FING_ON)	ONSET0_C0_FING_z
target word consonant onset (<u>M</u> ima)		ONSET1_C1_FING_z
target word lexically stressed vowel onset (M <u>i</u> ma)		ONSET2_V1_FING_z
pitch accent onset		ONSET3_PA_FING_z
f0 contour onset		ONSET4_F0_FING_z
pre-boundary consonant target (<u>m</u> a Mima)	pointing gesture target (FING_PEAK)	TARGET0_C0_FING_z
target word consonant target (<u>M</u> ima)		TARGET1_C1_FING_z
target word lexically stressed vowel target (M <u>i</u> ma)		TARGET2_V1_FING_z
intensity peak		TARGET3_INT_FING_z
pitch accent peak		TARGET4_PA_FING_z
f0 contour peak		TARGET5_F0_FING

3.2.2.1.2 Contour clustering

During prosodic verification, different types of prosodic contours on the target word were identified. These include (1) a H pitch accent in the first syllable followed by a fall and a low plateau on the second syllable, (2) a L pitch accent in the first syllable (with a flat f0 contour) followed by a rise in the second syllable, (3) a late H pitch accent that begins in the first syllable and reaches its peak in the second syllable (often the onset of the second syllable), followed by a high plateau or a fall. To explore how these different contours might affect coordination patterns,

f0 contour clustering was performed on the target word for the emerging clusters to be entered into statistical analysis described in the following subsection. The clustering method of Kaland (2023, 2026) was used with normalized f0 (z-score standardized by participant) to account for speaker variability.

3.2.3 Statistical analysis

The intervals described in **Table 5** were z-scored by interval type and speaker to account for speaker differences and were subsequently pooled. Extreme outliers (those exceeding three times the interquartile range) were removed after inspection (n=4). Statistical analyses were carried out in R (R Core Team, 2025).

To establish the most stable interval, that is, the coordination landmarks (RQ1), the variability of intervals was quantified and the differences between these variabilities were assessed. Participant-controlled variances were calculated to isolate within-participant variability, accounting for differences in baseline performance across participants. To control for remaining participant-specific effects, residual variance per interval was estimated from a linear mixed-effects model with participant-level random intercepts (*lme4* package, Bates, Maechler, Bolker, & Walker, 2015). Residual variances were then rescaled relative to the least variable interval (such that the least variable interval was assigned a value of 1). To assess if the differences in the variances of the intervals are statistically significant, bootstrapping with replacement (Davidson & Hinkley, 1997) was implemented, using the *boot* package (Canty & Ripley, 2021). Bootstrapping is useful in working with small sample sizes with variable distribution patterns (see e.g. Chernick (1999)'s guide and Egbert & Plonsky (2021) for an overview in the context of linguistics). The method produced a variance distribution for each

interval with 95% confidence intervals computed. Differences between the intervals' variances are considered significant if the confidence intervals do not include the reference level (=1).

To examine if coordination is more stable in pragmatically more prominent conditions (RQ2), the effect of information structure condition on the most stable interval (as ascertained in the previous step) was estimated. A linear mixed-effects model was fitted with *IS_SUBTYPE* (for information structure categories) as fixed effect, *PARTICIPANT* as random effect and the least variable interval as dependent variable. As participants produced distinct intonation patterns, the f0 contour clusters were entered into the analysis as levels of a *CONTOUR* factor along with the interaction of *CONTOUR* and *IS_SUBTYPE* to estimate their effect on variability, with *p*-value estimation via *Satterthwaite's method* using the *lme4* package (Bates et al., 2015). Significance is set at *p*<0.05. Note that the exploratory question of the role of a pre-target word boundary tone (*ma#Mima*) in coordination could not be examined as participants did not produce boundary tones consistently on the pre-target word, and four participants did not produce boundary tones altogether.

Information structure and f0 contour model:

$$Y = \beta_0 + b_0 + \beta_1(\text{IS_SUBTYPE}) + \beta_2(\text{CONTOUR}) + \beta_3(\text{IS_SUBTYPE} \times \text{CONTOUR}) + \varepsilon$$

The following R packages were used for data processing and visualization: *readr* (Wickham, Hester, & Bryan, 2023), *readxl* (Wickham & Bryan, 2023), *dplyr* (Wickham, François, Henry, Müller, & Vaughan, 2023), *tidyr* (Wickham, Vaughan, & Girlich, 2023), *ggplot2* (Wickham, 2016), *ggsignif* (Ahlmann-Eltze & Patil, 2021), *stringr* (Wickham, 2022), *ggpubr* (Kassambara, 2023).

3.3 Results

3.3.1 Smallest variance of speech and co-speech gesture intervals [RQ1]

The boxplots and overlaid scatterplots in **Figure 9** below visualize the z-score standardized temporal distances of the speech-co-speech gesture intervals. Boxplots 1-5 are onset-to-onset intervals, and boxplots 6-11 are target-to-target intervals. Visual inspection reveals that the intervals are similar in variability. **Table 6** lists the intervals' residual variances in ascending order, after controlling for participant variation, showing that the interval with the smallest variance is TARGET2__V1_FING_z.

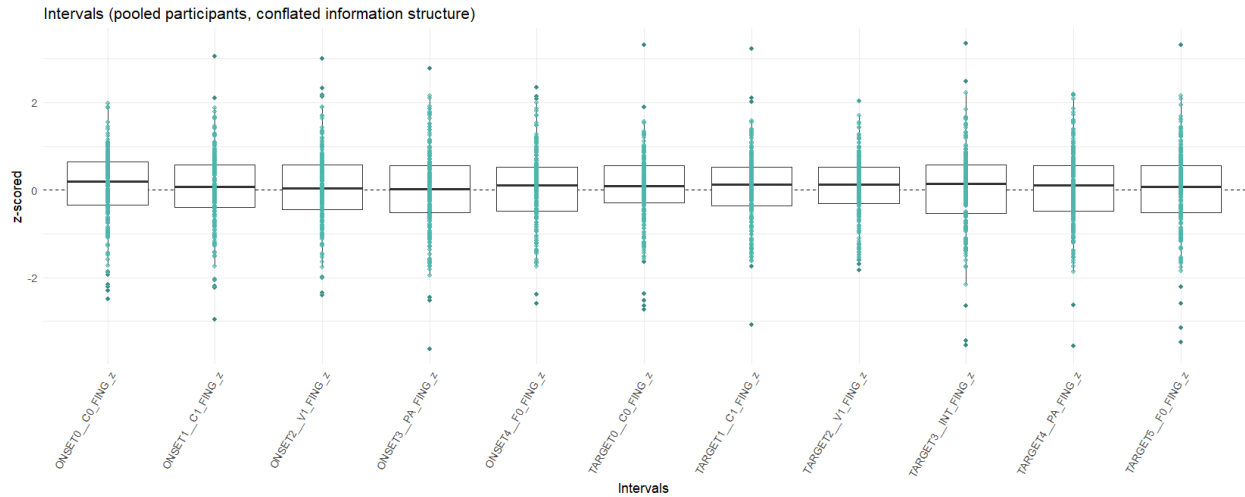


Figure 9. Boxplots and scatterplots of each interval (standardized)

Table 6. Residual variances of the intervals.

Interval	Residual Variance
TARGET2__V1_FING_z	0.544
TARGET1__C1_FING_z	0.649
ONSET0__C0_FING_z	0.656
TARGET0__C0_FING_z	0.665
ONSET4__F0_FING_z	0.704
ONSET1__C1_FING_z	0.705
ONSET2__V1_FING_z	0.706

Interval	Residual Variance
TARGET4__PA_FING_z	0.767
TARGET5__F0_FING_z	0.840
TARGET3__INT_FING_z	0.854
ONSET3__PA_FING_z	0.897

Evaluating the significance of the differences in variance, results of bootstrapping are presented in **Table 7**. The interval with the least variability, TARGET2_V1_FING_z, is set as the baseline (relative variance = 1), and the variance ratios of the remaining intervals are computed relative to this baseline. Bootstrap 95% confidence intervals (CIs) indicate that two intervals do not differ significantly from the reference, as their CIs include 1: ONSET0__C0_FING_z (pre-target word-initial consonant onset to pointing onset) and ONSET4_F0_FING_z (f0 onset to pointing onset). The CIs corresponding to the relative variance of the remaining intervals do not include the baseline of 1, showing significantly greater variance than the reference interval. To indicate the magnitude of the differences, **Figure 10** visualizes the relative differences across the speech-co-speech gesture intervals, including the computed confidence intervals.

Table 7. Relative variance ratios relative to TARGET2__V1_FING_z (the baseline group) and results of bootstrapping (confidence intervals). Nonsignificant results are shaded light gray.

Interval	Relative variance ratio	Lower CI	Upper CI
TARGET2__V1_FING_z	1	N/A	N/A
TARGET1__C1_FING_z	1.193	1.032	1.328
ONSET0__C0_FING_z	1.205	0.778	1.795
TARGET0__C0_FING_z	1.222	1.005	1.564
ONSET4__F0_FING_z	1.293	0.996	1.642
ONSET1__C1_FING_z	1.295	1.001	1.661
ONSET2__V1_FING_z	1.296	1.015	1.648
TARGET4__PA_FING_z	1.408	1.086	1.861
TARGET5__F0_FING_z	1.543	1.221	2.060
TARGET3__INT_FING_z	1.569	1.250	2.109

Interval	Relative variance ratio	Lower CI	Upper CI
ONSET3__PA_FING_z	1.647	1.158	2.795

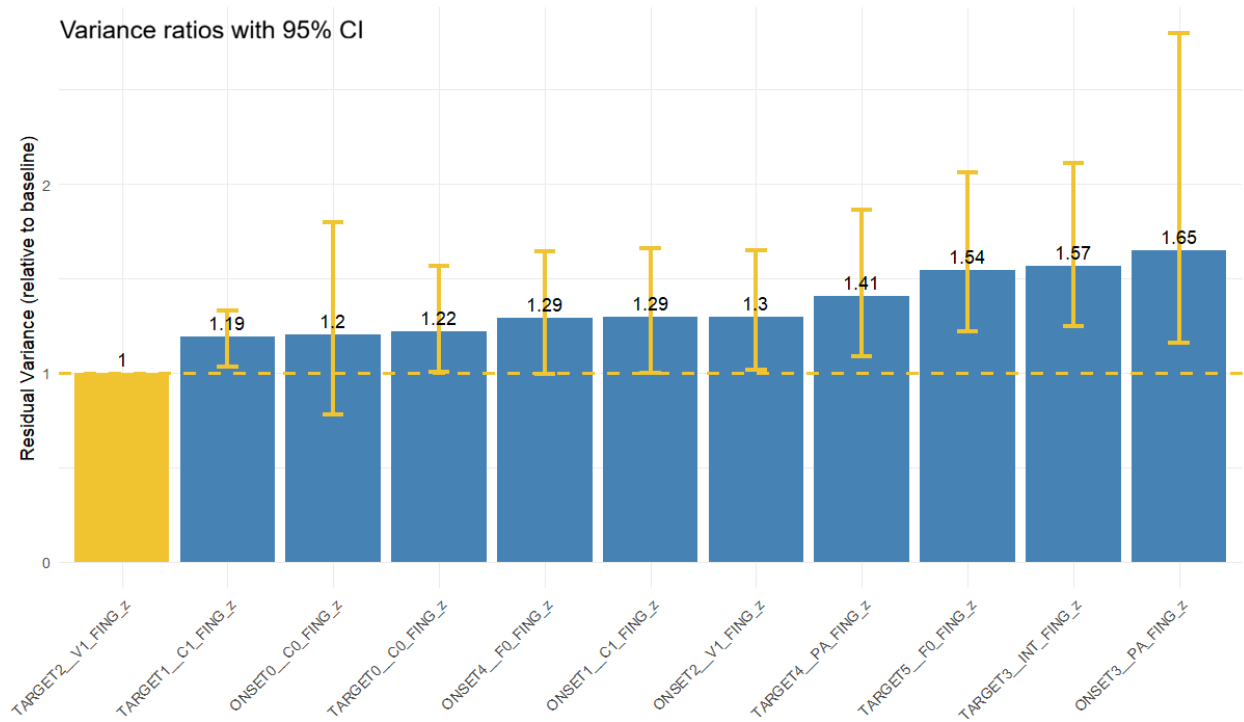


Figure 10. Variance ratios with bootstrap 95% confidence intervals relative to the interval with the smallest variance (TARGET2__V1_FING_z).

The figure shows that intervals with greater relative variance ranging from 1.41 to 1.65 (TARGET4__PA_FING_z, TARGET5__F0_FING_z, TARGET__INT_FING_z, ONSET3__PA_FING_z) differ robustly from the reference group, and intervals with smaller relative variance either do not reach significance (as listed above) or reach significance by a smaller margin.

To gauge if the data is driven by individual differences (i.e. speakers using different strategies for coordination), **Table 8** summarizes for each participant the raw variances and median values (negative values indicate that the pointing landmark follows the speech landmark). The table shows that of the relevant intervals, TARGET2__V1_FING is the least variable for all speakers, and it is the one closest to zero for two speakers (where 0 means perfect

temporal alignment between the pointing and speech landmarks). Given this overall pattern, TARGET2__V1_FING will be considered the least variable interval for the remaining analyses on information structure and f0 contour effects on coordination.

Table 8. By-participant raw median (ms) and variance (ms) of the intervals between the statistically not significantly different speech and co-speech gesture landmarks.

	ONSET0__C0_FING		ONSET4__F0_FING		TARGET2__V1_FING	
	median	variance	median	variance	median	variance
P1	139	341	480	289	-117	222
P3	-218	205	168	305	-219	54
P4	-571	297	-227	236	-405	184
P5	-600	1241	35.5	276	-108	151
P6	-144	155	95.3	134	-202	65
P7	-120	89	146	82	-80.4	81

3.3.2 Effects on coordination [RQ2]

3.3.2.1 Effect of information structure

The boxplots and overlaid scatterplots in **Figure 11** visualize the most stable interval (TARGET2__V1_FING_z) across the information structure subtypes. The interquartile ranges are ordered from smallest to greatest, which illustrates that the variability of the interval forms a gradient, and in the corrective focus, contrastive topic, and aboutness topic conditions, it appears to have smaller variability than in the broad and confirmation focus conditions. However, the linear mixed-effects model reveals no significant main effect of *IS_SUBTYPE* on the TARGET2__V1_FING_z value, indicating that none of the information structure types differ significantly from the reference level in their effect on the most stable interval. Given the lack of a significant effect of information structure, the following questions were not explored further:

(1) Is coordination more stable when the target word is in focus than when it is part of the

background? (2) Does the stability of coordination vary as a function of pragmatic prominence, meaning greater stability under greater pragmatic prominence than under less prominence? (3) Is coordination more stable in narrow focus (which is syntactically marked in Hungarian) than in broad focus (which is syntactically unmarked in Hungarian)?

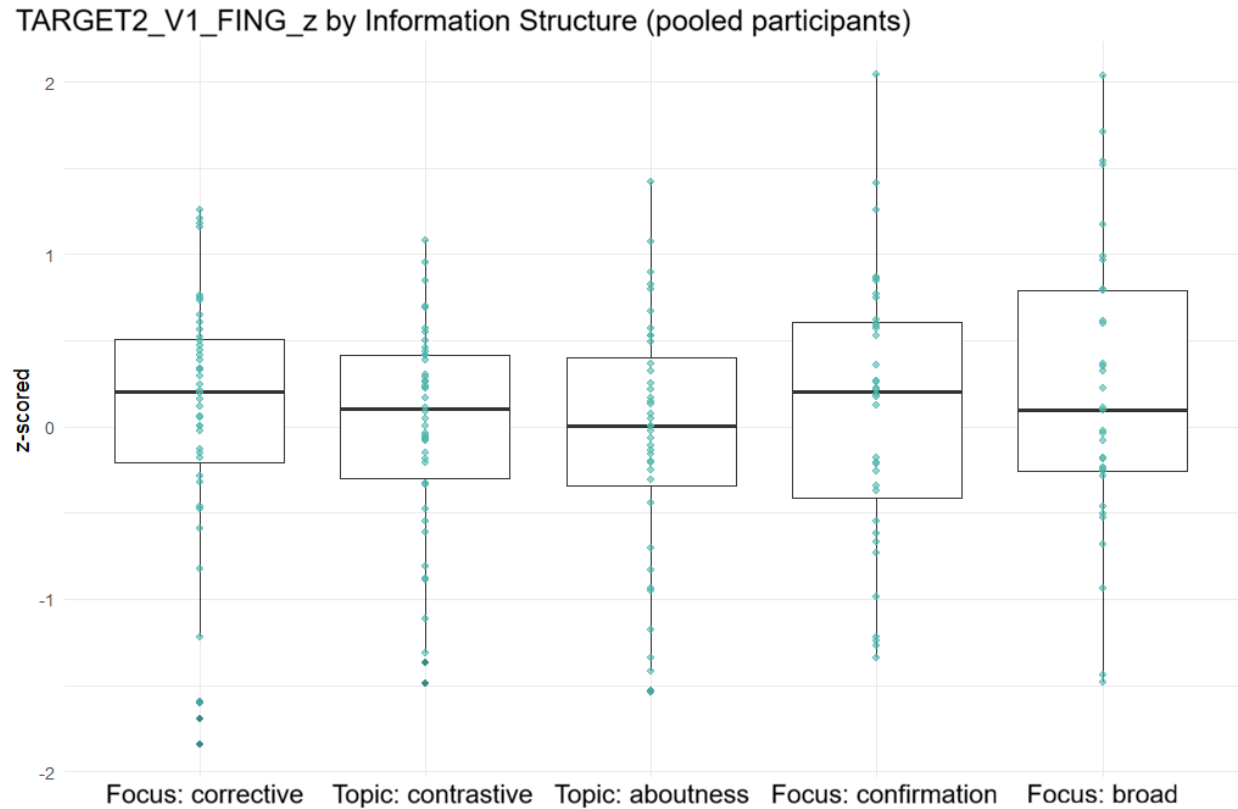


Figure 11. The most stable interval (TARGET2__V1_FING_z) across information structure conditions, ordered from smallest to greatest IQR.

3.3.2.2 Effect of *f0* contour type

As detailed in the introduction, intonation has a significant role in coordination in other languages; therefore, its effect on coordination was examined for Hungarian as well. **Figure 12** illustrates the *f0* contours speakers produced on the target word *Mima* [mimo]. Cluster 1 contains

contours with an early f0 peak, followed by a fall, cluster 2 contains contours with a low f0 on the first syllable followed by a rise in the second syllable, cluster 3 contains contours with a rising first syllable and a high plateau in the second syllable. Next, **Figure 13** plots the most stable interval (TARGET2__V1_FING_z) across the clusters. The figure shows that across the contour types, the IQRs have similar widths, demonstrating stability. The linear mixed-effects model underscores this observation, revealing no significant main effect of *CONTOUR* on the TARGET2__V1_FING_z interval, indicating that none of the contour types differ significantly from the reference level.

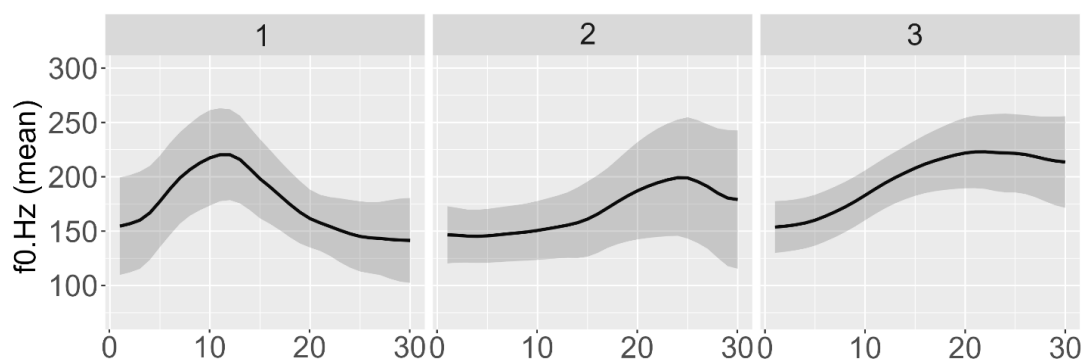


Figure 12. Target word f0 contour clusters.

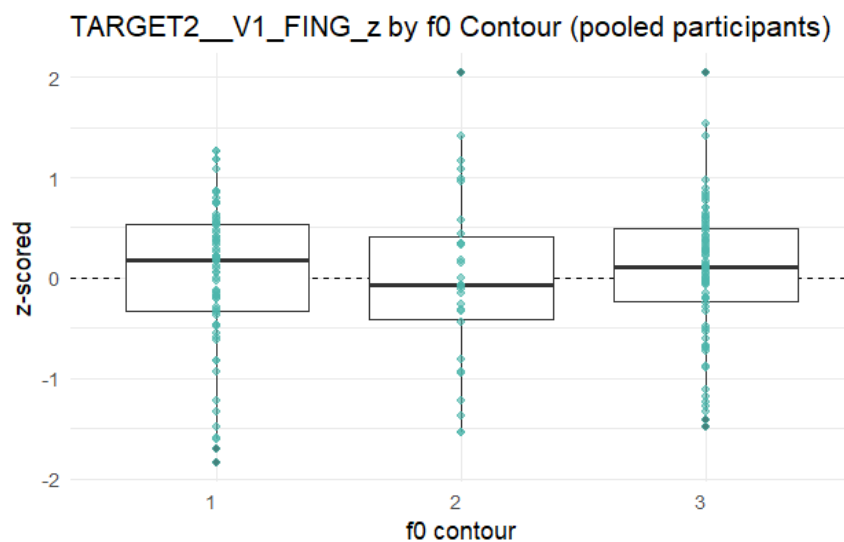


Figure 13. The most stable interval (TARGET2__V1_FING_z) across F0 contour clusters ordered from smallest to greatest IQR.

3.3.2.3 Interaction of information structure and f0 contour type

To examine whether information structure and prosody jointly influence coordination, the interactional effect of the two was estimated. The number of tokens in each f0 contour cluster is given for each information structure type in **Table 9** (the table enumerates the tokens that could be included in contour clustering and in the coordination analysis as well, since tokens where the visual and EMA data could not be matched for gesture labeling were necessarily excluded). The table shows that broad focus tokens occur in all clusters, but mainly group into cluster 3 (72%), confirmation focus tokens group mainly into clusters 1 (59.37%) and 3 (34.37%), corrective focus predominantly groups into cluster 1 (82.92%), aboutness topic tokens group mainly into clusters 2 (31.57%) and 3 (50%), and contrastive topic tokens predominantly group into cluster 3 (63.63%). Cluster 1 consists mainly of narrow focus tokens, cluster 2 of topics, and cluster 3 has many tokens from all information structure categories, except corrective focus.

Table 9. Number of tokens from information structure conditions in each f0 contour cluster.

Information Structure	f0 contour cluster		
	1	2	3
Focus: broad	5 (20%)	2 (8%)	18 (72%)
Focus: confirmation	19 (59.37%)	2 (6.25%)	11 (34.37%)
Focus: corrective	34 (82.92%)	5 (12.19%)	2 (4.87%)
Topic: aboutness	7 (18.42%)	12 (31.57%)	19 (50%)
Topic: contrastive	7 (15.9%)	9 (20.45%)	28 (63.63%)

The boxplots and overlaid scatterplots in **Figure 14** visualize the most stable interval (TARGET2__V1_FING_z) across information structure conditions and all f0 contour clusters. Although the IQRs show variability across the board, the linear-mixed effects model reveals no significant main effects and interaction between IS_SUBTYPE and CONTOUR, indicating that information structure and intonation do not significantly affect coordination.

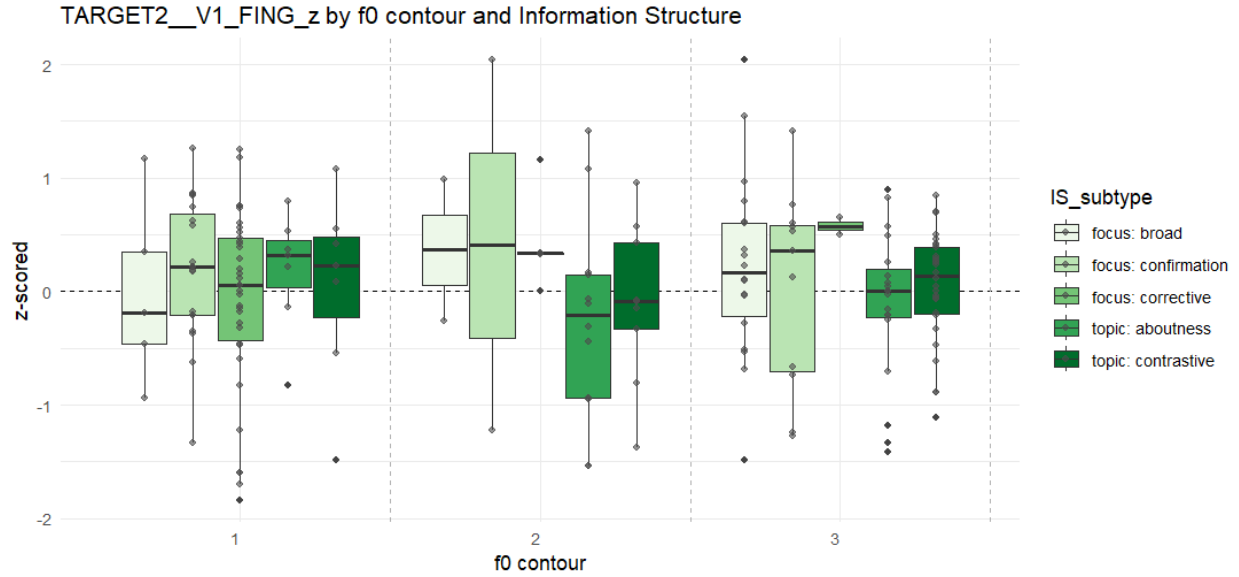


Figure 14. The most stable interval (TARGET2__V1_FING_z) across information structure conditions and f0 contour clusters.

3.4 Discussion

3.4.1 Coordination [RQ1]

3.4.1.1 Coordinating landmarks in Hungarian

Answering the first research question, the study tests which aspect of speech gestures (constriction or prosodic) coordinate with which aspect of pointing gestures in Hungarian, focusing on onsets and targets of gestures in each modality and applying the concept of *gestural coordination* as implemented in Articulatory Phonology (Goldstein, Byrd, & Saltzman, 2006; Browman & Goldstein, 1995; Browman & Goldstein, 1992). Among conceivable landmarks of speech and co-speech gestures, the pair of landmarks with the least variable time interval (lag) between them is considered to have the most stable temporal relationship, and so to constitute the points of coordination. A number of onset-to-onset and target-to-target intervals were tested,

with onsets and targets of constriction (C, V) and prosodic (pitch accent, f0, intensity) gestures paired with the onset and target of the pointing gesture.

The interval that emerged as most stable was the one between the articulatory target of the first vowel in the target word (*Mima*) and the pointing gesture target (*TARGET2__V1_FING_z*, referred to as vowel target – pointing target interval moving forward), indicating a target-to-target coordination pattern in Hungarian. As lexical stress in Hungarian falls on the first syllable of content words (Varga, 2002; Siptár & Törkenczy, 2000; Kálmán & Nádasdy, 1994), the landmark of coordination, the first vowel, is the nucleus of the stressed syllable. Whether this coordination is due to (a) lexical stress as a prominence-lending property or (b) the first syllable as the left-edge (or onset) of the target phrase attracting the co-speech gesture cannot be disambiguated in the present study given the stimuli, and in any case, the two generally coincide in Hungarian. What can be concluded is that the pointing gesture target coordinates with the lexically stressed syllable, which is on the left edge of the lexical affiliate (i.e. the word the pointing gesture corresponds to, which in the present case is the target word *Mima*). A brief discussion of the other examined intervals follows.

The *vowel target – pointing target* interval is significantly less variable than every other target-to-target interval considered for the analysis. As potential prosodic anchoring points, pitch accent targets, f0 contour maxima, and intensity peaks were included in the analysis, and all three intervals are significantly more variable than the *vowel target – pointing target* interval, indicating a lack of evidence for pitch accents, f0, and intensity having an organizing role in coordination with pointing gestures.

The *vowel target – pointing target* interval is also significantly less variable than most onset-to-onset intervals (see **Table 7**), with the exception of the onset of the pre-target word (*ma*

“today”; ONSET0__C0_FING_z) and the onset of the labeled f0 contour in the target word (ONSET4__F0_FING_z). Note that although the differences are not statistically significant, these two intervals have greater variances than the *vowel target – pointing target* interval (as also shown in **Figure 10**) and considering speakers’ raw data separately (**Table 8**), the pattern holds with the latter interval being the least variable. As a secondary measure, the speaker-specific median distance of the landmarks was also shown for the three intervals; in two speakers’ production, the *vowel target – pointing target* interval had the smallest lag, and in the other four speakers’ production, the *f0 onset – pointing onset* interval was shortest (note that these differences were not statistically evaluated).

The overall pattern of data supports the conclusion that the nucleus of the first, lexically stressed syllable of the lexical affiliate coordinates with the pointing gesture target. Future work will have to confirm or contradict the present assessment.

3.4.1.2 Contextualization

Prior studies provide ample evidence that prosodic structure plays a significant role in speech and co-speech gesture alignment. In particular, prosodically prominent units of speech have been shown to systematically co-occur with co-speech gestures in languages that express prosodic prominence via pitch accentuation (with or without additional syntactic requirements). Studies find co-speech gestures to co-occur with pitch accented syllables (Kügler & Gregori, 2023; Türk & Calhoun, 2023; Rohrer et al., 2023; Shattuck-Hufnagel & Ren, 2018; Esteve-Gibert & Prieto, 2013; Loehr, 2012; Leonard & Cummins, 2011; Mendoza-Denton & Jannedy, 2011; Roustan & Dohen, 2010; Swerts & Krahmer, 2010; Yasinnik et al., 2004), and the gestural target of the co-speech gesture, that is, the apex, to align with the pitch accent (Kügler & Gregori, 2023; Türk & Calhoun, 2023; Leonard & Cummins, 2011; Roustan & Dohen, 2010;

Loehr, 2012; Mendoza-Denton & Jannedy, 2011; Yasinnik et al., 2004), and when a pitch accent is absent, the co-speech gesture may align with an accentless tone on the edge of a prosodic word, as in Turkish (Türk and Calhoun, 2023).

A smaller number of studies report on alignment in languages that do not express prominence via pitch accentuation. In Cantonese, co-speech gestures align with the first syllable in disyllabic words (Fung & Mok, 2018), and as described in a conference talk, they align in particular with the vowel onset in both mono- and disyllabic words (conference talk: Rohrer, Hong, & Bosker, 2026, *DGfS*). Similarly, in Mandarin, beat gestures align with the vowel onset of monosyllabic (CV) lexical affiliates (Rohrer, Hong, & Bosker, 2024). Broadly, the co-speech gesture appears to be anchored to the vowel of the first syllable in the target word in both languages.

Cross-linguistic differences between tone languages are reported in Franich, Keupdjio, & Nwosu (2025), who find that in Medumba, the word stem-initial position anchors co-speech gestures, while in Igbo, the word-final syllable does so, with the metrical structure of the tonal foot playing a role in this alignment. Further, Franich (2026) reports bilingual speakers of Babanki and Cameroonian English to align speech and co-speech gestures differently, depending on the language being used. Altogether, studies thus show cross-linguistic differences which may relate to the prosodic typology of the languages examined.

The data presented in this study aligns to some extent with the cross-linguistic observations in that we find coordination between the target of the pointing gesture and the vowel of the target-initial syllable, which carries prominence in Hungarian. As also mentioned above, we cannot distinguish between coordination with the syllable due to it being stressed and prominence-bearing (as in the first set of studies) and due to its initial position in the target word

(as in Cantonese and Mandarin), because these co-occur in the language. However, lexical stress is a predictable prosodic feature of the language (always occurring on the first syllable), and as such, its nucleus is not an unlikely anchor. Overall, coordination with the target-initial stressed syllable is consistent with prosodic structure playing a role in speech and co-speech gesture coordination in Hungarian.

Where Hungarian differs is observable due to the examination of articulatory data. While we do find, broadly speaking, that the target of the pointing gesture coordinates with the stressed syllable, we also find that it is neither the associated pitch accent nor the onset of the vowel that anchors the co-speech gesture; instead, we find coordination with the vowel target. How much of this difference may be due to prosodic typology (with alignment being mediated through the prosodic system of the given language), the concept and measure of alignment, the genre of speech, and/or the elicitation method may remain to be seen in future research. Within the theoretical framework that the present study adopted, i.e. Articulatory Phonology (AP), however, an argument can be made for a typological distinction.

As noted previously, alignment is conceptualized in the present study as *gestural coordination*, with the stability of the temporal relationship between gestures forming its basis. This concept is originally modeled in AP for the coordination of gestures composing consonants and vowels forming a syllable, but it has been applied to the modeling of lexical tones coordination with constriction gestures in Mandarin (Gao, 2008), and it has additionally been extended to include the temporal relationship of post-lexical tones (intonation) with constriction gestures in pitch accented syllables (Mücke et al., 2012). Most pertinent to the discussion, the concept of gestural coordination has been applied to the temporal relationship of speech and co-speech gestures as well in Krivokapić et al. (2016). The study examines the coordination of

constriction gestures, pitch accents, and pointing gestures in English and finds the most stable temporal relationship to be between the pitch accent target and the pointing gesture target. The present study is similar in methodology to Krivokapić et al. (2016), and the findings are also similar inasmuch as a target-to-target coordination pattern is identified in both cases, but they differ in the speech landmark that anchors the co-speech gesture (pitch accent vs. vowel target), indicating a degree of departure from prior observations, and suggesting that typological, rather than methodological factors are at play in the cross-linguistic difference found.

The discussion of Hungarian moves forward, in the present chapter (as well as in **Chapter 5**, which discusses the relationship between prosody and co-speech gestures in more detail, informed also by the findings reported in **Chapter 4**) with the interpretation that prosodic structure influences gestural coordination in language specific ways, and in Hungarian the (*target-word initial, lexically stressed-*) *vowel target – pointing target* interval is indeed the relevant one for coordination.

3.4.2 Effects on coordination in Hungarian [RQ2]

3.4.2.1 Information structure

The second research question probes the role of information structure in speech and co-speech gestural coordination, asking if the source of pragmatic prominence and relative prominence have an effect on the stability of coordination, with most stability under focus-related and/or greater prominence. Supporting evidence for a focus-related effect on temporal alignment was found in German (Gregori & Kügler, 2026; Kügler & Gregori, 2023), where alignment between the speech and co-speech landmarks, specifically pitch accent peaks and co-speech gesture apices, was tighter, i.e. with the landmarks occurring closer together in focus than

in background. The authors attribute the phenomena to hyperarticulation due to focus, extending the theory of Hypo- and Hyperarticulation (Lindblom, 1990) from speech articulation to the co-speech gestural modality.

In the present study, five information structural conditions were included in the analysis: broad focus as a baseline, two types of narrow focus (corrective and confirmation), and two types of topics (aboutness and contrastive). When part of a broad focus sentence, the target word is syntactically unmarked and does not have additional prominence relative to the rest of the sentence. In contrast to this, the narrow focus and topic types are syntactically marked (that is, the target word takes a different syntactic focus or topic position in the left periphery; see the introductory chapter for a discussion of this as well as the relevant references). Further, they differ in relative pragmatic prominence, with corrective focus and contrastive topic having additional pragmatic functions relative to confirmation focus and aboutness topic, respectively.

Results do not show evidence for a modulating effect of information structure on speech and co-speech gesture coordination in Hungarian, as measured in the stability of the temporal relationship between the vowel target of the first syllable and the pointing gesture target (i.e. the least variable interval). As such, the contrasts between broad vs. narrow focus, focus vs. topic, and relative pragmatic prominence were not explored further statistically. The null results indicate that, unlike in German (Gregori & Kügler, 2026; Kügler & Gregori, 2023), hyperarticulation due to focus may not be present in Hungarian, at least not in the temporal aspect of gesture production. Whether other aspects of co-speech gesture realization may be subject to hyperarticulation remains to be explored.

3.4.2.2 *F0 contour and information structure*

The study also examined the potential effect on coordination of f0 contours on the target word and their interaction with information structure. The primary motivation in including contour type as a potential factor in the analysis is the possibility of participants producing distinct focus and topic contours (H* on the stressed syllable followed by a fall, or L* on the first syllable followed by a rise, respectively), the relative prominence of which may influence coordination. Given the *vowel target* – *pointing target* coordination and the relative stability of fixed stress in Hungarian, an interaction effect was not particularly likely (especially taking the lack of significant main effects of information structure and f0 contour into consideration), but the possibility was explored nonetheless, as there is cross-linguistic evidence of relatively greater prominence affecting temporal alignment (see Carignan et al. 2024 for an interactional effect in head nod strategies in French).

Three contour clusters were identified: (1) falling with an early high peak on the first syllable followed by a fall, (2) rising with low f0 on the first syllable followed by a rise in the second syllable, and (3) rising with an earlier rise followed by a high plateau in the second syllable. There is no one-to-one correspondence between the contours and information structure, but there is a tendency for the target word in some of the categories to group into one or two contour clusters. In particular, 82.92% of the tokens in the corrective focus condition were produced with an early peak followed by a fall (contour cluster 1). 59.37% of the confirmation focus tokens were also produced with this contour, but 34.37% have instead a later peak followed by high plateau (contour cluster 3). Both the broad focus and the contrastive topic tokens tended to group into cluster 3 as well (72% and 63.63%, respectively), as did 50% of the

aboutness topics, with another 31.57% being produced with a low f0 followed by a rise in the second syllable (contour cluster 2).

With regards to an effect on the least variable interval (*vowel target – pointing target*), statistical modeling showed that neither the shape of the contour nor its interaction with information structure have a significant effect, indicating that the interval is not modulated by these factors.

3.4.2.3 Contextualization

How these results relate to prior studies of other languages is a complex question due to methodological and conceptual differences. Since the present study involves elicited co-speech gestures, the results cannot be situated in the existing body of work on the *frequency* of co-occurrence of information structure types and gestures. Instead, the focus is on studies that examine the characteristics of the temporal lag between landmarks (as a measure of stability of coordination or tightness of alignment) as a function of information structure.

In particular, Kügler & Gregori (2023, also Gregori & Kügler, 2026) find an effect of focus on the tightness of speech and co-speech gesture alignment when compared with background in German, which the authors interpret as prominence-related hyperarticulation in the gestural modality. Inasmuch as we can perform valid cross-study comparisons, and assuming that language-specific points of coordination can be compared, as far as statistically significant results go, our study does not align with the observations about the German data. We find no overall effect of information structure and see no categorical differences in coordination between our focus (broad/information, narrow-confirmation, narrow-corrective) and background (aboutness and contrastive topic) conditions in the Hungarian data.

3.5 Conclusion

3.5.1 Broader implications

Beyond providing a first description of speech and co-speech gesture coordination in Hungarian, the present findings have broader implications for our understanding of coordination across languages. In contrast to temporal alignment patterns reported for other languages (among them English, German, French, Catalan, and Turkish) that employ pitch accents to mark prominence primarily or in addition to syntactic marking, in Hungarian, pitch accent, f_0 , and intensity peaks do not appear to serve as temporal anchors for pointing gesture targets. Instead, our results suggest that pointing gestures are coordinated with the vowel of the first, lexically stressed syllable of the lexical affiliate. These findings, as well as those on tone languages, indicate that the nature of speech and co-speech gesture coordination is shaped by the prosodic structure of the language in question. The findings therefore underscore the importance of typologically informed research with more studies employing similar methodologies to facilitate cross-linguistic comparisons and the development of a comprehensive theory of multimodal gesture coordination.

3.5.2 Limitations and future research avenues

A number of limitations should be acknowledged. First, due to this being a kinematic study, the sample size is relatively small, which requires caution in the interpretation of the findings and in their generalizability. As an initial investigation of speech and co-speech gesture coordination in Hungarian, however, the study provides a valuable foundation for future research. Second, the data was collected via a controlled experiment and due to the high number of repetitions, there is a possibility that the elicitation method resulted in some task-specific

behavior. However, to our knowledge, there is no evidence of a qualitative and theoretically relevant difference in speech and co-speech gesture coordination between genres of speech, and a clear benefit of this method is that it allows for the examination of co-speech gesture behavior in specific theoretically meaningful conditions. Follow-up studies will need to address the issue of level of spontaneity and genre of speech.

Next, Franich, Keupdjio, & Nwosu (2025) suggest that the function pitch has in a language may modulate gestural alignment. Considering the relative roles of syntax and prosody in prominence marking, trade-off effects have been noted in other languages (see e.g. Zou & Arnhold, 2026), and recent work on Hungarian has shown as well that prosody may take on a greater functional load in marking focus either when syntactic marking reaches its limit or when syntactic movement is not performed. To the first point, Langer & K  gler (2023) find prosodic marking of contrastive focus within a larger syntactically marked narrow focus phrase. For instance, given a question “Did Bence spot the yellow snake?”, the entire DP is moved to focus position in the answer “No, he spotted [the *red* snake]!” in Hungarian, but only the adjective *red* carries contrast, which can be marked via intonation modulation. In our study, the target word constitutes the entire focus (and topic) phrase, therefore, prosody does not have an additional disambiguating function. To the second point, Buza et al. (2026) show in a perception study with Hungarian listeners that prosodic focus marking can be an acceptable alternative to syntactic focus marking when the context clues are present. The question arises whether intonational landmarks (pitch accent, f0, boundary tones) factor into speech and co-speech gesture coordination in Hungarian in contexts in which prosody does function as a marker of contrast within narrow focus.

Another future avenue of research, and one that mitigates some of the limitations of the present study, concerns non-manual gestures. The data collection method of the present study allows for the analysis of further co-speech gestures in the future, including eyebrow, head, and to some extent, torso movement. Given the experimental methodology, these gestures constitute a spontaneously produced set; therefore, patterns of both the frequency of co-occurrence and the stability of coordination (as defined in Articulatory Phonology) can be assessed in relation to information structure and prosody.

Chapter 4 Boundary and Prominence Effects

4.1 Introduction

The previous chapter examined the nature of speech and co-speech gesture coordination in Hungarian, and the present chapter explores the effect of prosodic boundaries and pragmatic prominence on the temporal properties of pointing gestures in an effort to better understand the extent to which co-speech gestures and linguistic structures are intertwined.

As explained in the introductory chapter, *prosodic boundaries* are the edges of prosodic phrases of various levels in the prosodic hierarchy. Boundaries are marked by certain temporal properties, with articulatory gestures being longer, larger, and less overlapped at phrasal boundaries. These boundary-related lengthening effects are modeled in Articulatory Phonology as being due to a temporal modulation gesture, called the π -gesture (Byrd & Saltzman, 2003). The π -gesture model predicts that the amount of lengthening increases with *boundary strength* (for reviews, see Paschen et al., 2022; Byrd & Krivokapić, 2021; Krivokapić, 2014; Fletcher, 2010). For instance, given a boundary related to an intermediate phrase (ip) vs. one related to an Intonation Phrase (IP), the IP boundary is stronger and its effect greater, producing a greater amount of lengthening in the boundary-adjacent segments relative to lengthening related to the ip-boundary. As illustrated in **Figure 15**, boundary-related lengthening is present at either end of a boundary; that is, phrase-initially as well as phrase-finally, and the effect is strongest where the π -gesture has the strongest activation and decreases as activation decreases (see Byrd & Krivokapić, 2021), indicated by shading in **Figure 15**.

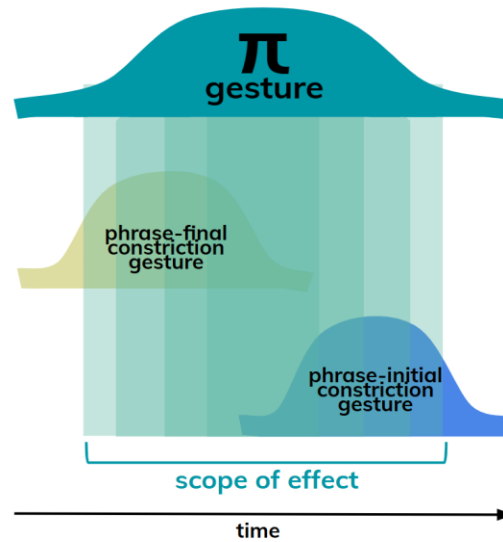


Figure 15. π -gesture effect scope. Figure adapted from Byrd & Krivokapić (2021)

The close alignment between speech and co-speech gestures gives rise to the hypothesis that the temporal properties of not only constriction gestures but also of co-speech gestures may be affected by prosodic structure. In support of this, as surveyed in the introductory chapter, an utterance initial strengthening effect on gestures was found in Kim et al. (2014), and, in a small study with two speakers, prosodic boundaries have been shown to lengthen speech and co-speech gestures alike in English, as reported in Krivokapić et al. (2017). As boundary-related lengthening has been reported for Hungarian (e.g. Gósy & Krepsz, 2018; White & Mády, 2008), similar lengthening effects as found in Krivokapić et al. (2017) may apply to co-speech gestures as well.

In addition to boundary-related lengthening, Krivokapić et al. (2017), Rusiewicz et al (2014), and Parrell et al (2014) find prominence-related lengthening of co-speech gestures. Lengthening due to prominence is modeled in Articulatory Phonology as the effect of a temporal modulation gesture called the μ -gesture (Saltzman et al., 2008). The effect of prominence is expected to be the strongest at the syllable nucleus and taper out away from it (Bombien et al.,

2013). There is some evidence for prominence having a modulating effect on vowel duration in the prominent syllable in Hungarian as well (Markó et al., 2018) although Vogel, Athanasopoulou, & Pincus (2015) and White & Mády (2008) report evidence to the contrary. In addition to the possibility of lengthening, the insertion of pauses (or: of prosodic boundaries) has been noted as a possible marker of prominence in Hungarian (Mády, Kleber, Reichel, & Szalontai, 2016). The present chapter thus investigates these questions related to the temporal characteristic of manual gestures applied to Hungarian. The specific research questions and hypotheses are listed below.

[RQ3] How do prosodic boundaries affect co-speech gestures?

Hypothesis: If co-speech gestures are affected by prosodic boundaries (i.e., the π -gesture), a lengthening effect on parts of the co-speech gesture that overlap with a prosodic boundary will be observable.

[RQ4] How does prominence affect the temporal properties of co-speech gestures?

Hypothesis: If prominence affects the temporal properties of speech and co-speech gestures, information structure corresponding to greater pragmatic prominence will have a lengthening effect (corrective focus, contrastive topic > broad focus; corrective focus > contrastive topic).

[RQ5] How does the interaction of prosodic boundaries and prominence influence the temporal properties of co-speech gestures?

Hypothesis: If prosodic boundaries and prominence both have a lengthening effect on co-speech gestures, a cumulative lengthening effect will be observable with more lengthening under a prosodic boundary given greater prominence than given smaller prominence.

The overarching research question is how the temporal properties of co-speech gestures may be affected by prosodic boundaries, prominence, and their interplay (again, focusing on

pointing gestures), with the goal of examining the interrelationship of prosodic structure and co-speech gestures in Hungarian, providing cross-linguistic insight into the prosody-gesture interface.

4.2 Methodology

The data collection procedure is as described in the methodology chapter. The stimuli materials, data analysis, and statistical methods relevant to the present chapter are detailed below.

4.2.1 Stimuli

To examine the effect of a *preceding/initial* boundary and prominence on the temporal properties of speech and co-speech gestures, the following materials were used. Question and answer pairs were constructed so as to elicit three levels of prominence (stemming from information structural differences) on the target word “Mima” in two boundary conditions in the answer. The prominence types included were (1) broad focus, (2) corrective focus, (3) contrastive topic. The boundary types included were (a) word boundary and (b) IP boundary. Modulating the prosodic boundary that precedes the target word, the lexical item *ma* (“today”) that immediately precedes the target functions as a topic in the word boundary condition and as an echo question in the IP boundary condition.

[1] Broad focus: “What happened at the competition today?”

a. Word boundary: ***Ma Mima** bebetonozta a hírnevét.*

“Today, Mima cemented her reputation.”

b. IP boundary: ***Ma? Mima** bebetonozta a hírnevét.*

“Today? Mima cemented her reputation.”

[2] Corrective focus: “Was Papi exceptionally good today?”

- a. Word boundary: ***Ma Mima** betonozta be a hírnevét.*
“Today, Mima cemented her reputation.”
- b. IP boundary: ***Ma? Mima** betonozta be a hírnevét.*
“Today? Mima cemented her reputation.”

[3] Contrastive topic: “What happened today at the competition with you and Mima?”

- a. Word boundary: ***Ma Mima** bebetonozta a hírnevét. Én pedig csak csodáltam a sok ügyes munkát.*
“Today, Mima cemented her reputation. And I just admired all the skillful work.”
- b. IP boundary: ***Ma? Mima** bebetonozta a hírnevét. Én pedig csak csodáltam a sok ügyes munkát.*
“Today? Mima cemented her reputation. And I just admired all the skillful work.”

The materials and experimental structure address the hypotheses related to prosodic boundaries and prominence under controlled conditions, as further detailed in the methodology chapter.

4.2.2 Data analysis

4.2.2.1 Data labeling

As noted in the previous chapter, the target word was followed by the *bebetonozta* (verbal prefix – verb) sequence in the broad focus and contrastive topic conditions, and it was followed by the sequence *betonozta be* (verb – verbal prefix) in the corrective focus condition. In all cases, therefore, the phonetic sequence was [mɒ mimɒ be]. The pre-target word consonant (C0), the target word consonant and vowels (C1, V1, C2, V2), and the post-target word consonant (C3) were labeled. The constriction and pointing gestures were labeled in *Mview*

(Mark Tiede, Yale University), using the *FindGest* method as described in the methodology chapter. In the case of two speakers, the pointing gestures, due to their shape, were labeled using the *snapex* method in *Mview*; for this reason, only the onset and target measurement points are available. In all other cases, the gesture onset (ON), peak velocity (to target; PVEL), nucleus onset (NONS), maximum constriction (TARGET), nucleus offset (NOFFS), peak velocity (to offset, PVEL2), and gesture offset (GOFFS) labels are available (see **Figure 19** for a schematic representation).

4.2.2.2 Boundaries

The magnitude of the pre-target juncture (word vs. IP boundary) can be measured in the duration of the pause between the words involved. The longer the pause, the bigger the prosodic boundary, and – if the pointing gesture is influenced by prosodic structure – the greater the lengthening. To confirm that participants produced the intended boundaries (word and IP), the pause preceding the target word (*ma_?_Mima*) was measured as the lag between the maximum constriction of the pre-target onset (C0 target) and the target onset (C1 target) (see Byrd & Saltzman, 1998). **Figure 16** below plots the raw values across speakers and boundary types.

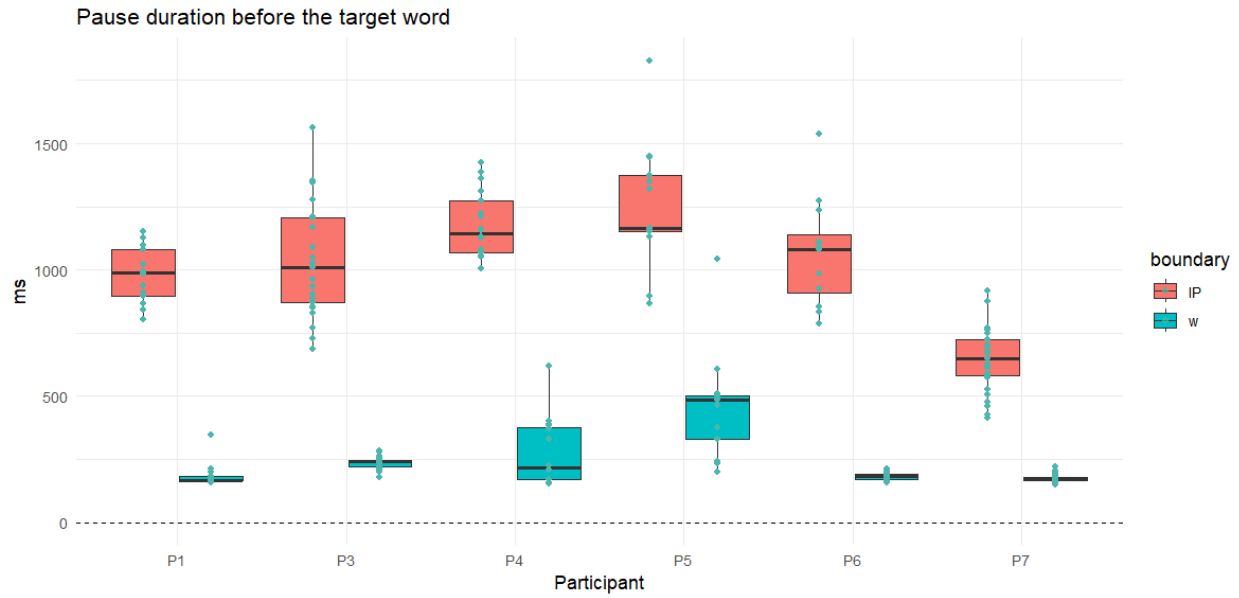


Figure 16. Raw values of pause duration before the target word, separated by boundary type and participant

As the figure shows, pauses preceding the target word are greater for all speakers in the IP condition than in the word condition, indicating that participants maintained a distinction between the targeted boundaries.

In addition to producing pauses before the target word, participants frequently paused between the target word and the following word (*Mima__be-*), producing a prosodic boundary. The effect of the inserted boundary on the pointing gesture can be examined in a post-hoc exploratory analysis. The expected effect is lengthening of the pointing gesture that is directly proportional to the duration of the post-target pause. Pause duration was measured as the lag between the maximum constriction of the last consonant of the target word (C2 target) and the first consonant of the post-target word (C3 target). The raw values of pause durations are plotted in **Figure 17**.

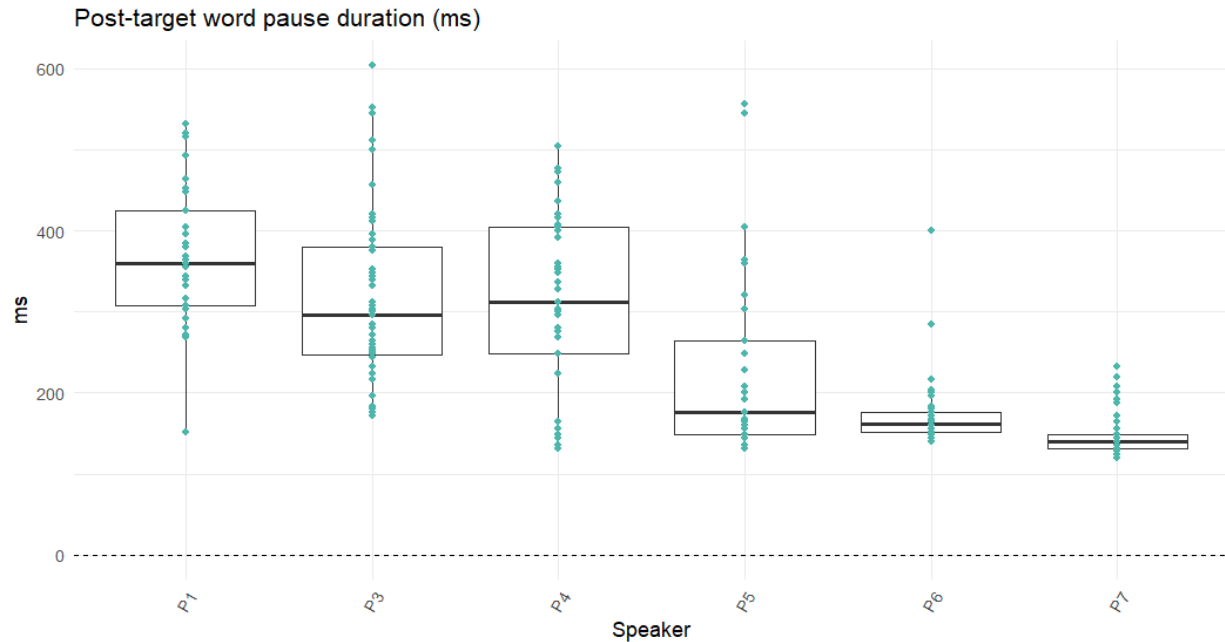


Figure 17. Raw values of pause duration after the target word, separated by participant.

4.2.2.3 Co-speech gesture position relative to target

As explained in the introduction, the effect of prosodic boundaries on constriction gestures is strongest near the boundary; for pointing gestures, an effect may be expected on phases of the gesture that either overlap with or are close enough to the boundary. To establish where, relative to the target word and relative to the boundaries, the co-speech gestural phases are realized, the intervals listed in **Table 10** were calculated. In the case of constriction gestures, we calculated the intervals including the onset or target of landmarks, and in the case of the pointing gestures, intervals included the onset, plateau onset, maximum constriction, plateau offset, or gesture offset landmarks (where available). In each combination, the timepoint of the pointing gesture landmark was subtracted from the timepoint of the articulatory landmark.

Table 10. Measured lags for gauging gesture position relative to the target word

Lag coding	Articulatory landmark	Pointing landmark
	M A #/IP M I M A # B... C0 C1 V1 C2 V2 C3	
<i>s1_P_on_c1_on</i>	C1 onset	Gesture onset
<i>s2_P_nons_v2_max</i>	V2 target	Plateau onset
<i>s3_P_target_c3_max</i>	C3 target	Target
<i>s4_P_noffs_c3_max</i>	C3 target	Plateau offset
<i>s5_P_off_c3_max</i>	C3 target	Gesture offset

The raw data of the measured lags are visualized in **Figure 18**, with participants and boundary conditions separated. As the figure indicates, considering participants separately, the position of the pointing gesture is rather similar across boundary conditions, with the notable difference that the co-speech gesture onset precedes the target word onset in the word condition but lags behind it in the IP condition (*s1_P_on_c1_on*) for participant P3 and P7. Similar tendencies can be observed for P4 and P5, with the gesture onset lagging behind the target word onset in both boundary conditions, but with a tendency for an earlier gesture onset in the word boundary condition. In the case of P1, the pointing gesture onset occurs earlier than the target word onset in both boundary conditions, with more variability in the IP condition. For P6, the pointing onset and target word onset occur closer together. Overall, the pointing onset appears to be earlier in the word than in the IP boundary condition.

Regarding the position of the pointing plateau, as the *s2_P_nons_V2_max* interval shows, its onset tends to occur around the last segment of the target, and its offset occurs after the target of the post-target word onset (C3 = *ma Mima le*) for P3, P6, and P7, and before the C3 target for P1 (as the *s4_P_noffs_c3_max* interval shows). Finally, as the *s5_P_off_c3_max* interval shows, the pointing gesture offset occurs after the C3 target for P1, P3, P6, and P7. These intervals are not available for P4 and P5. The pointing target occurs between the pointing plateau onset and offset; it precedes C3 in P5's production and occurs at or after C3 in P4's.

Overall, the raw data indicates that in both boundary conditions, the co-speech gesture onset to plateau region temporally aligns with the target word (with a tendency for a later onset in the IP boundary condition), the pointing onset to target region overlaps with the target word and the post-target boundary, the pointing plateau begins during the last segment of the target word and lasts into the post-target verb, and the offset of the pointing gesture occurs later.

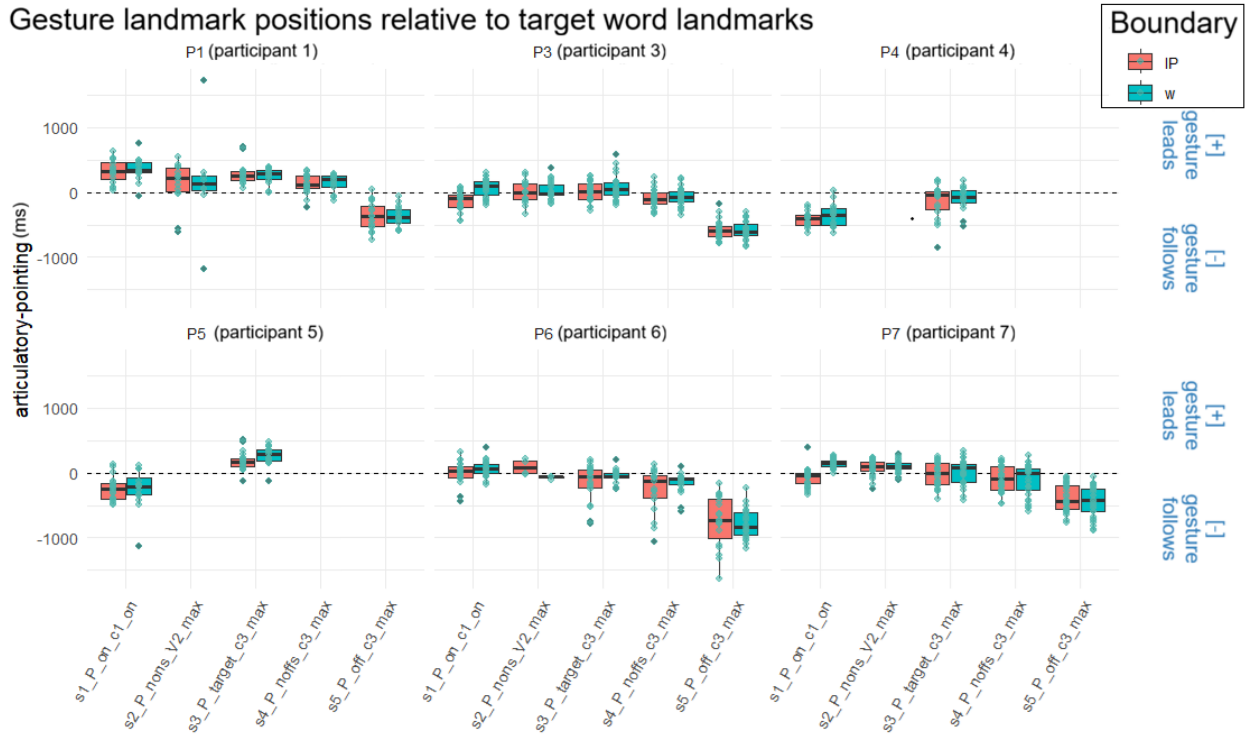


Figure 18. Co-speech gesture position relative to the target word, separated by participant and boundary condition. The co-speech gestural timepoints were subtracted from the articulatory landmarks' timepoints (in ms). Negative range indicates that the co-speech gestural landmark follows the articulatory landmark.

The gesture onset-to-target phase may not overlap with the initial pause, but some lengthening effect of the boundary (i.e. of the π -gesture) may be detectable as the effect of the boundary tapers out away from the boundary (as is the case with constriction gestures in general, and as shown on the schematic representation in **Figure 15** above). Additionally, the pointing plateau overlaps with the juncture between the target word and the following word, therefore the

post-hoc analysis may show a lengthening effect of the post-target boundary on this gesture phase. The target-to-return phase is unlikely to show lengthening due to its relatively late position.

4.2.2.4 Boundary and prominence related measures

A schematic representation of speech and co-speech gesture landmarks is given in **Figure 19**. To examine the effects of prosodic boundaries and prominence, the durations of the speech and co-speech gestural phases were measured. The phases are listed in **Table 11** and shown in **Figure 20**.

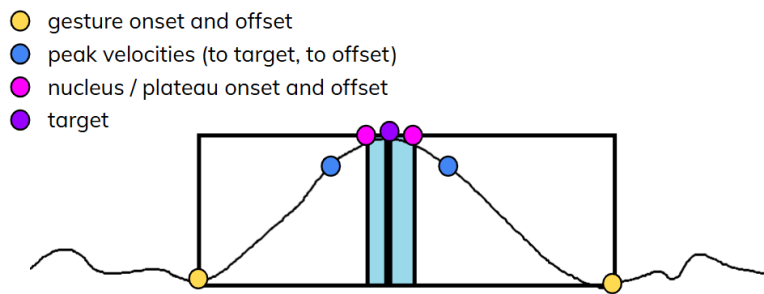


Figure 19. Schematic representation of gesture landmarks.

The initial boundary is expected to show an effect on C1 (the onset of the target word); the measures included are the duration of the lip closing movement (*a1_c1_forming*), the duration of the plateau (*a2_c1_plat*), and the duration of the lip opening movement (*a3_c1_releasing*). The boundary may show an effect on V1 (the first vowel of the target word); the corresponding measures included are the duration of the tongue raising movement from [ɐ] to [i] (*a4_v1_forming*), the duration of the plateau (*a5_v1_plat*), and the duration of the tongue lowering movement from [i] to [ɐ] (*a6_v1_releasing*). The expectation is that these durations are longer in the IP-boundary condition than in the word-boundary condition, as well as in

information structure conditions that correspond to greater prominence. (Note that the plateau for constriction gestures is included for future reference, as segment duration in Hungarian is phonemic for both vowels and consonant, and the related durational differences may be expected in the plateau.)

Given where the pointing gesture is situated relative to the target word, the boundaries, and the nucleus of the stressed syllable, any effect of the initial boundary and of prominence is to be expected on the onset to target interval (*p1_point_dur*). Effects of a boundary between the target word and the post-target word may be expected on the plateau of pointing (*p2_point_plat*), as well as on the return phase (*p3_point_re*). Note that due to the unavailability of the measures involved, the latter two phases cannot be examined for participants P4 and P5.

Table 11. Speech and co-speech gestural landmarks. (Note that the prefixes in the interval coding scheme (*p1_*, *p2_*, *a1_*, *a2_*, etc.) are not meaningful.)

Gesture phase durations		From	To	Interval coding
pointing phase	pointing	pointing onset	pointing target	<i>p1_point_dur</i>
	pointing plateau	NONS	NOFFS	<i>p2_point_plat</i>
	pointing return	pointing target	pointing offset	<i>p3_point_re_dur</i>
C1 [mimɒ]	closing	C1 onset	C1 target	<i>a1_c1_forming</i>
	closing plateau	C1 NONS	C1 NOFFS	<i>a2_c1_plat</i>
	opening	C1 target	C1 offset	<i>a3_c1_releasing</i>
V1 [mimɒ]	raising	V1 onset	V1 target	<i>a4_v1_forming</i>
	vowel plateau	V1 NONS	V1 NOFFS	<i>a5_v1_plat</i>
	lowering	V1 target	V1 offset	<i>a6_v1_releasing</i>

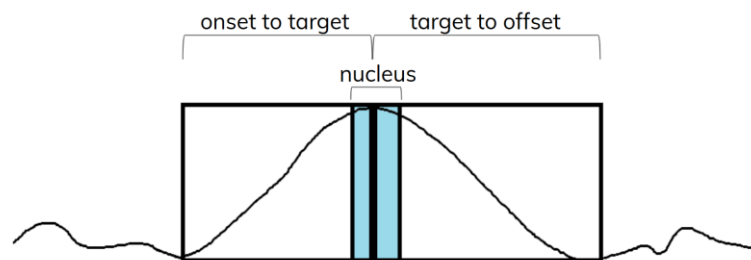


Figure 20. Speech and pointing gestural regions of interest.

4.2.3 Statistical analysis

Raw duration values were filtered to exclude extreme values over 2000 ms (n=1). To account for speaker variability, duration values were z-score standardized within each participant and interval (as listed in **Table 11**). All statistical analyses were conducted in R (R Core Team, 2025).

To examine initial-boundary related lengthening (RQ3), by-participant linear models were fitted (*lm* function). The primary predictor of interest was *BOUNDARY* with two levels (word vs. IP) and the dependent variable was the z-scored interval duration. To examine articulator-specific effects, separate models were fitted for each interval (lip aperture intervals: *a1_c1_forming*, *a2_c1_plat*, *a3_c1_releasing*), tongue dorsum intervals: *a4_v1_forming*, *a5_v1_plat*, *a6_v1_releasing*, and pointing finger intervals: *p1_point_dur*, *p2_point_plat*, *p3_point_re*).

Initial-boundary model:

$$Y^{INTERVAL, SPEAKER} = \beta_0 + \beta_1(BOUNDARY) + \epsilon$$

To assess the effect of prominence (RQ4), inferred from information structure, by-participant linear models (*lm* function) were fitted. As in the boundary-model, separate models were fitted for each speech and co-speech gesture interval. The dependent variable is the z-scored interval duration; the predictor is *IS_SUBTYPE* (for information structure). In addition to constriction gestures, the effect of information structure on the duration of the pre-target pause is examined. By-participant linear models were fitted for each participant with normalized pause duration (z-scored by participant) as the dependent variable and *IS_SUBTYPE* as predictor. Pairwise comparisons between information structure types were performed for each interval

using the model-estimated marginal means to prevent bias of unbalanced data (*emmeans* package), p-values were adjusted for multiple comparisons (Tukey-adjustment).

Prominence model:

$$Y^{INTERVAL, SPEAKER} = \beta_0 + \beta_1(IS_SUBTYPE) + \varepsilon$$

To examine the interaction of boundary and prominence related effects (RQ5), by-speaker linear models were fitted (*lm* function). The dependent variable was the z-score normalized interval duration, the predictors were *BOUNDARY* and *IS_SUBTYPE*, and their interaction (*BOUNDARY* $\times$ *IS_SUBTYPE*). To estimate the significance of the interaction effect, pairwise comparisons of the estimated marginal means were computed for the interaction terms using the *contrast* function of the *emmeans* package; p-values were adjusted for multiple comparisons (Šidák-correction, Šidák, 1967).

Boundary and Prominence interaction model:

$$Y^{INTERVAL, SPEAKER} = \beta_0 + \beta_1(BOUNDARY) + \beta_2(IS_SUBTYPE) + \beta_3(BOUNDARY \times IS_SUBTYPE) + \varepsilon$$

Post-hoc analysis: As participants produced pauses after the target word, the effect of that boundary is also examined via by-participant linear models fitted for each pointing interval. The dependent variable was the z-scored interval duration, and the predictor was the post-target pause duration (*POST_PAUSE*).

Post-target boundary model:

$$Y^{SPEAKER} = \beta_0 + \beta_1(POST_PAUSE) + \varepsilon$$

Post-hoc analysis: To explore in further detail how the production of the post-target pause may relate to the targeted conditions of initial boundary and information structure, the

effects of *BOUNDARY* and *IS_SUBTYPE* on the pause duration (z-scored by participant) are assessed via by-participant linear regressions. Pairwise comparisons were conducted on the estimated model means using the *emmeans* package (Lenth R, 2023), *p*-values were adjusted for multiple comparisons (Tukey-adjustment).

Boundary effect on post-target pause duration:

$$Y^{SPEAKER} = \beta_0 + \beta_1(\text{BOUNDARY}) + \varepsilon$$

Prominence effect on post-target pause duration:

$$Y^{SPEAKER} = \beta_0 + \beta_1(\text{IS_SUBTYPE}) + \varepsilon$$

The R packages used for visualization were *ggplot2* (Wickham, 2016), *ggpubr* (Kassambara, 2023), *interactions* (Long, 2024), *ggsignif* (Ahlmann-Eltze & Patil, 2021), *stringr* (Wickham, 2022). Additional packages used for data handling were *readr* (Wickham, Hester, & Bryan, 2023), *readxl* (Wickham & Bryan, 2023), *dplyr* (Wickham, François, Henry, Müller, & Vaughan, 2023), and *tidyr* (Wickham, Vaughan, & Girlich, 2023).

4.3 Results

4.3.1 Initial-boundary effects [RQ3]

Initial-boundary effects were estimated via by-participant linear models fitted for each speech and co-speech gesture phase. **Figure 21** shows standardized durations of the lip aperture gesture intervals in C1 of the target word, separated by participant and initial-boundary type (IP vs. word boundary). The figure indicates that duration varies by boundary type with generally longer durations in the IP boundary condition.

The observation is supported by linear models, which show the following significant differences (see **Table 12** for the statistics). For participants P1, P3, P4, P6, and P7, the *a1_c1_forming* (consonant onset to maximum constriction) duration is significantly shorter in the word boundary condition. For participants P3, P6, and P7, the *a2_c1_plat* (consonant plateau) duration is significantly shorter in the word boundary condition, and a trend in this direction is observed for participant P4 as well. Lastly, for P3, P6, and P7, the *a3_c1_releasing* (consonant target to offset) duration is significantly shorter in the word boundary condition as well, and a trend in this direction is additionally observed for P1. No statistically significant difference in consonant movement durations can be observed in P5's production. With the exception of P5, the pattern shows significantly longer durations (or trends in this direction) of the onset consonant following an IP boundary.

Table 12. Linear model outputs (significant and marginal effects of boundary on C1)

<i>Participant</i>	<i>Estimate</i>	<i>Standard error</i>	<i>t-ratio</i>	<i>p-value</i>
<i>a1_c1_forming</i>				
P1	-1.2857	0.2823	-4.555	< 0.001
P3	-1.4203	0.1955	-7.263	< 0.001
P4	-1.3327	0.2546	-5.235	< 0.001
P6	-1.7654	0.1435	-12.3	< 0.001
P7	-1.4122	0.1876	-7.526	< 0.001
<i>a2_c1_plat</i>				
P3	-1.4510	0.1910	-7.598	< 0.001
P6	-1.0587	0.2655	-3.987	< 0.001
P7	-1.3835	0.1915	-7.225	< 0.001
P4	-0.6202	0.3271	-1.896	0.07
<i>a3_c1_releasing</i>				
P3	-1.1979	0.2233	-5.364	<0.001
P4	-0.8444	0.3115	-2.711	0.01
P6	-1.1995	0.2514	-4.728	<0.001
P7	-1.4467	0.1828	-7.913	<0.001
P1	-0.6344	0.3601	-1.762	0.09

Initial-boundary effect on C1 phases across participants

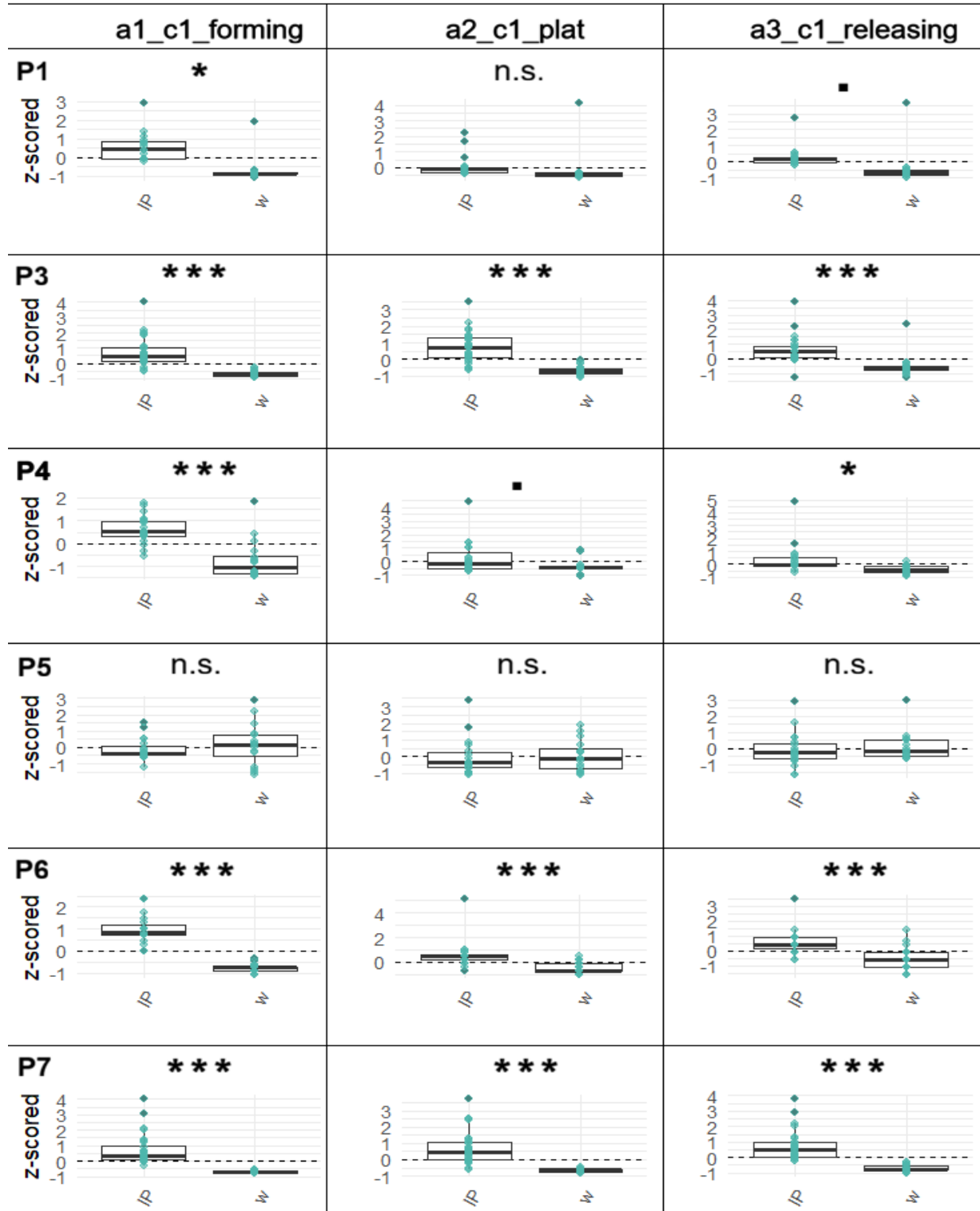


Figure 21. Initial-boundary effect on C1, separated by participant. Significance is indicated by asterisks, dots (.) indicate trends.

Considering the tongue dorsum gesture in V1 of the target word next, we find that the *a4_v1_forming* (vowel onset to target) durations are not significantly different in the production of any of the speakers, and the *a5_v1_plat* (vowel plateau) and *a6_v1_releasing* (vowel target to offset) durations differ significantly for two of the six speakers. The significant differences are shown in **Figure 22**, which plots the standardized durations of the tongue dorsum gesture intervals separated by participant and initial-boundary type (IP vs. word boundary). Linear models indicate for participants P4 and P7 that the *a5_v1_plat* duration is significantly shorter in the word boundary condition (P4: $\beta = -0.6906$, $SE = 0.3228$, $t = -2.140$, $p = 0.04$; P7: $\beta = -0.5968$, $SE = 0.2600$, $t = -2.295$, $p = 0.03$), and for participant P7, the *a6_v1_releasing* duration is significantly shorter in the word boundary condition as well ($\beta = -0.9175$, $SE = 0.2417$, $t = -3.795$, $p < 0.001$). No other vowel duration is significantly different in the production of any of the speakers. Overall, there are fewer durational differences for V1 across boundary conditions.

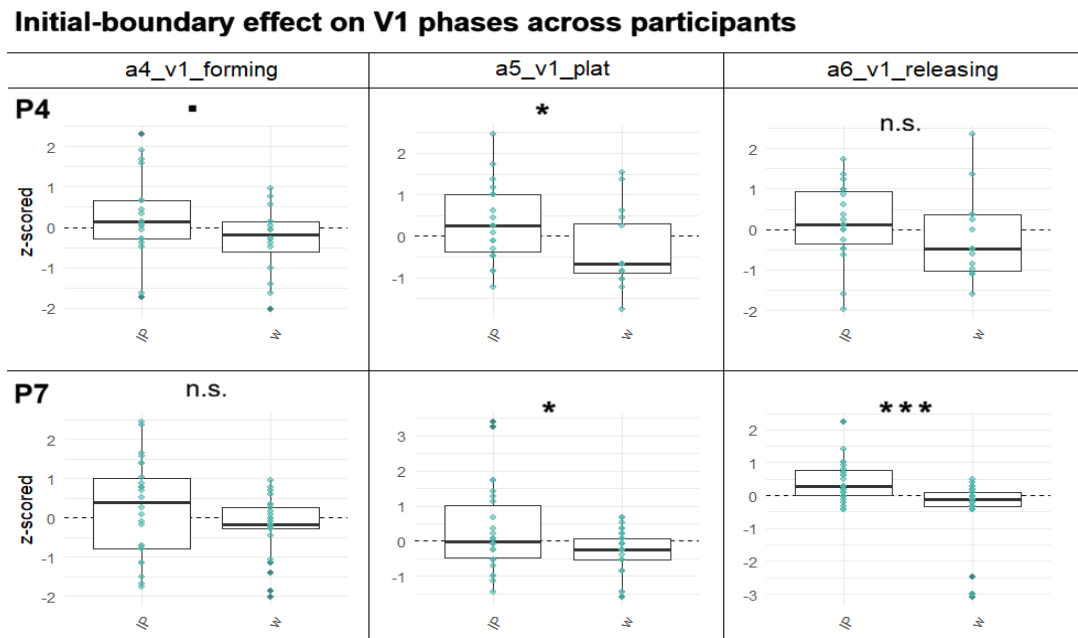


Figure 22. Initial-boundary effect on V1, separated by participant. Significance is indicated by asterisks, dots (.) indicate trends.

Turning to the pointing gesture next, few significant durational differences are observed. The *p1_point_dur* (pointing onset to target) duration is not significantly different in the production of any of the participants. **Figure 23** plots the standardized durations of the pointing gesture intervals, separated by initial-boundary type, for participants that differ significantly in at least one of the pointing phases. The figure shows a boundary effect on pointing for three speakers, such that the return movements are shorter in the IP boundary condition. Linear models support this observation. Specifically for participants P6, and P7, *the p3_point_re_dur* (pointing target to offset) duration is significantly longer in the word boundary condition (P6: $\beta = 0.7268$, $SE = 0.2921$, $t = 2.489$, $p = 0.01$; P7: $\beta = 0.7811$, $SE = 0.2457$, $t = 3.179$, $p = 0.002$), and a trend is observed for P3 in the same direction ($\beta = 0.5128$, $SE = 0.2709$, $t = 1.894$, $p < 0.06$). Additionally, the *p2_point_plat* (pointing plateau) duration is significantly longer in the word boundary condition for participant P3 ($\beta = 0.5994$, $SE = 0.2672$, $t = 2.243$, $p < 0.03$). No other pointing phase duration is significantly different across the boundary conditions.

Overall, significantly longer durations are observed on C1 durations in the IP boundary condition in the case of five of the six speakers, significantly longer durations of the V1 plateau and release durations in the IP boundary condition in two speakers' production, and significantly longer durations of the pointing plateau and return durations in the word boundary condition for three of the speakers.

Initial-boundary effect on pointing phases across participants

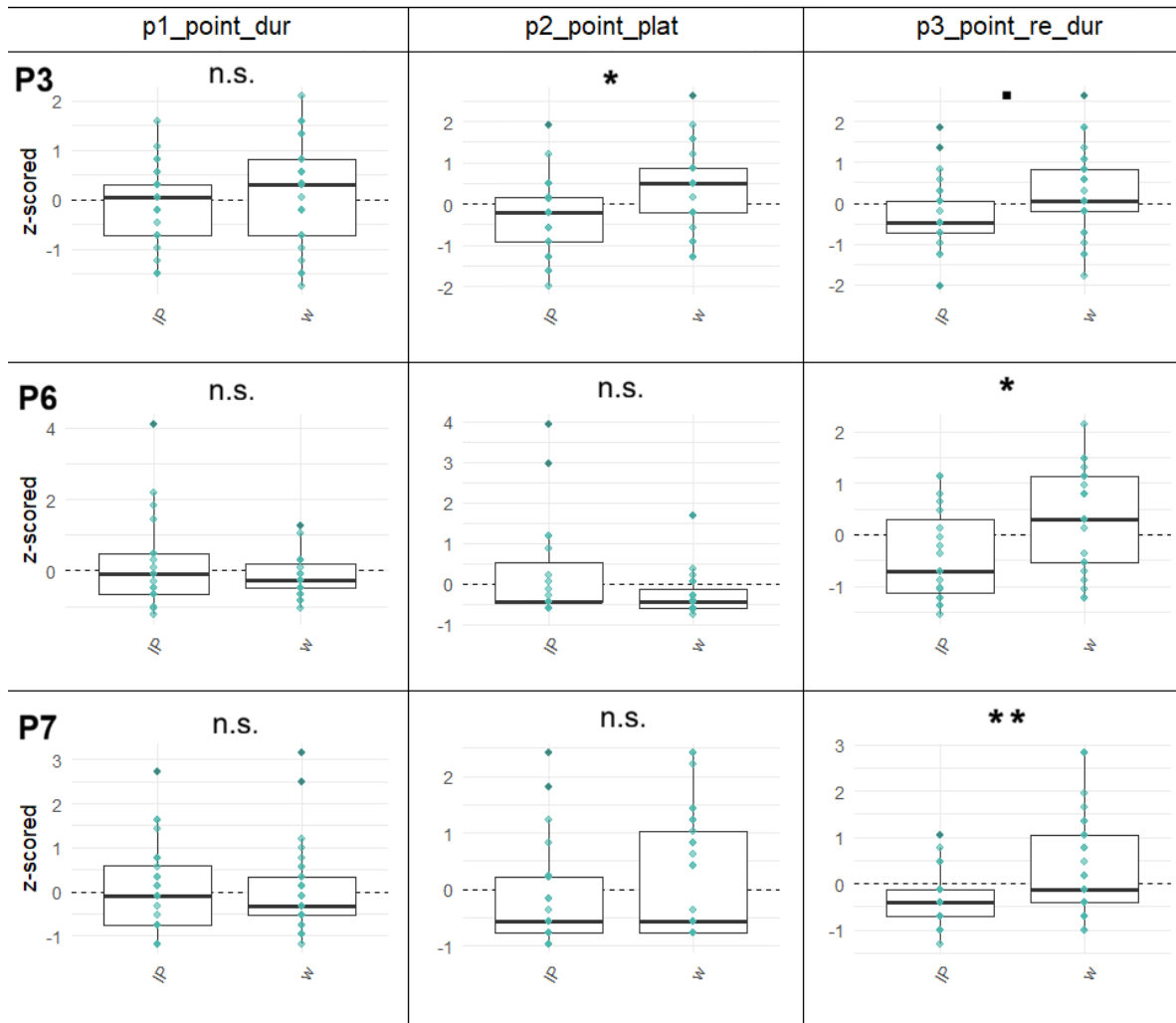


Figure 23. Initial-boundary effect on pointing phases, separated by participant. Significance is indicated by asterisks, dots (.) indicate trends.

4.3.2 Information structure effects [RQ4]

Information structure effects were estimated via by-participant linear models fitted for each speech and co-speech gesture interval, as well as for the pre-target pause; significant effects are examined via pairwise comparisons with p-value adjustment for multiple comparisons using the Tukey method. The models indicate no statistically significant effect on the consonant

durations for either speaker, as well as no significant effect on the vowel durations in the case of five speakers.

In participant P4's production, two significant lengthening effects were found (**Figure 24**). The duration of *a5_v1_plat* (vowel plateau) is significantly affected by information structure. Pairwise comparisons of model-estimated marginal means reveal a marginal lengthening effect of broad focus relative to corrective focus ($\beta = 0.959$, $SE = 0.442$, $t = 2.168$, $p = 0.09$). Additionally, in the same speaker's production, the duration of *a6_v1_releasing* (vowel target to offset) is significantly affected by information structure. Pairwise comparisons of model-estimated marginal means reveal a marginal lengthening effect of broad focus relative to corrective focus ($\beta = 1.012$, $SE = 0.434$, $t = 2.331$, $p = 0.07$). No evidence can be observed of further significant or marginal differences between speech gesture durations across IS conditions.

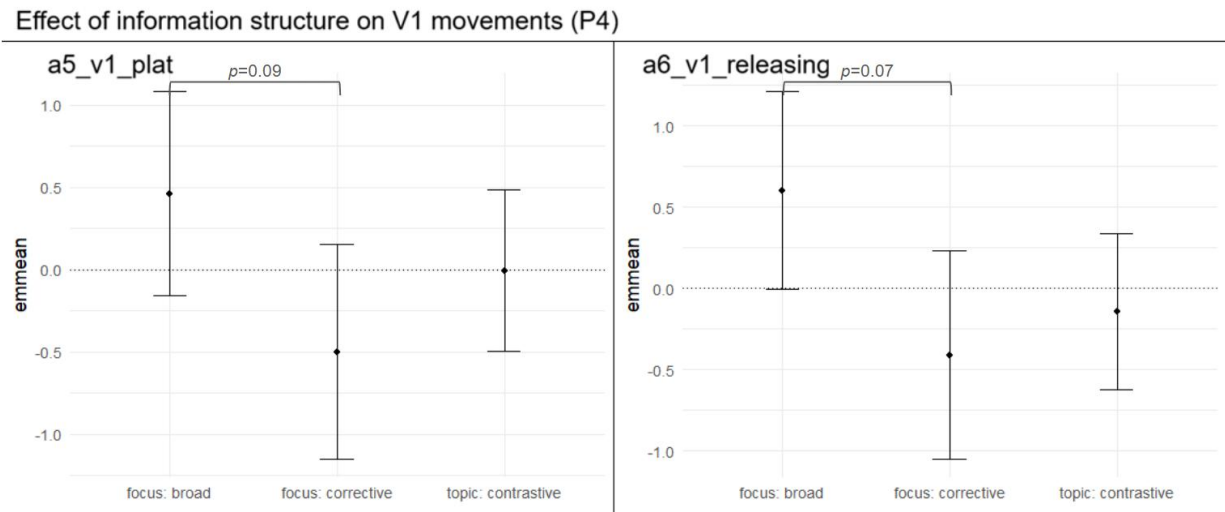


Figure 24. Marginal effect of information structure on two V1 movements (P4).

Regarding the effect of IS on pre-target pause duration, the following significant differences were observed. In the production of participants P1, P4, and P6, the pre-target pause

is significantly longer in broad focus than in corrective focus (**P1**: $\beta = 0.564$, $SE = 0.164$, $t = 3.448$, $p = 0.002$; **P4**: $\beta = 0.748$, $SE = 0.173$, $t = 4.322$, $p < 0.001$; **P6**: $\beta = 0.561$, $SE = 0.139$, $t = 4.035$, $p < 0.001$). In the production of P5 and P7, the opposite effect is found, with the pre-target pause being shorter in the broad focus condition than in corrective focus (**P5**: $\beta = -0.5556$, $SE = 0.204$, $t = -2.727$, $p = 0.02$; **P7**: $\beta = -0.3544$, $SE = 0.107$, $t = -3.315$, $p = 0.003$). In the same two speakers' production, the pause is significantly longer in the corrective focus condition than in contrastive topic as well (**P5**: $\beta = 0.4693$, $SE = 0.194$, $t = 2.414$, $p = 0.04$; **P7**: $\beta = 0.3394$, $SE = 0.108$, $t = 3.147$, $p = 0.005$). The opposite effect is found in the production of participant P6, with the pre-target pause being shorter in the corrective focus condition than in contrastive topic ($\beta = -0.423$, $SE = 0.133$, $t = -3.191$, $p = 0.005$). Finally, in the production of participants P3 and P4, the pre-target pause is significantly longer in broad focus than in contrastive topic (**P3**: $\beta = 0.313$, $SE = 0.120$, $t = 2.599$, $p = 0.03$; **P4**: $\beta = 0.427$, $SE = 0.147$, $t = 2.902$, $p = 0.01$). **Figure 25** illustrates significant contrasts.

Effect of information structure on initial pause duration

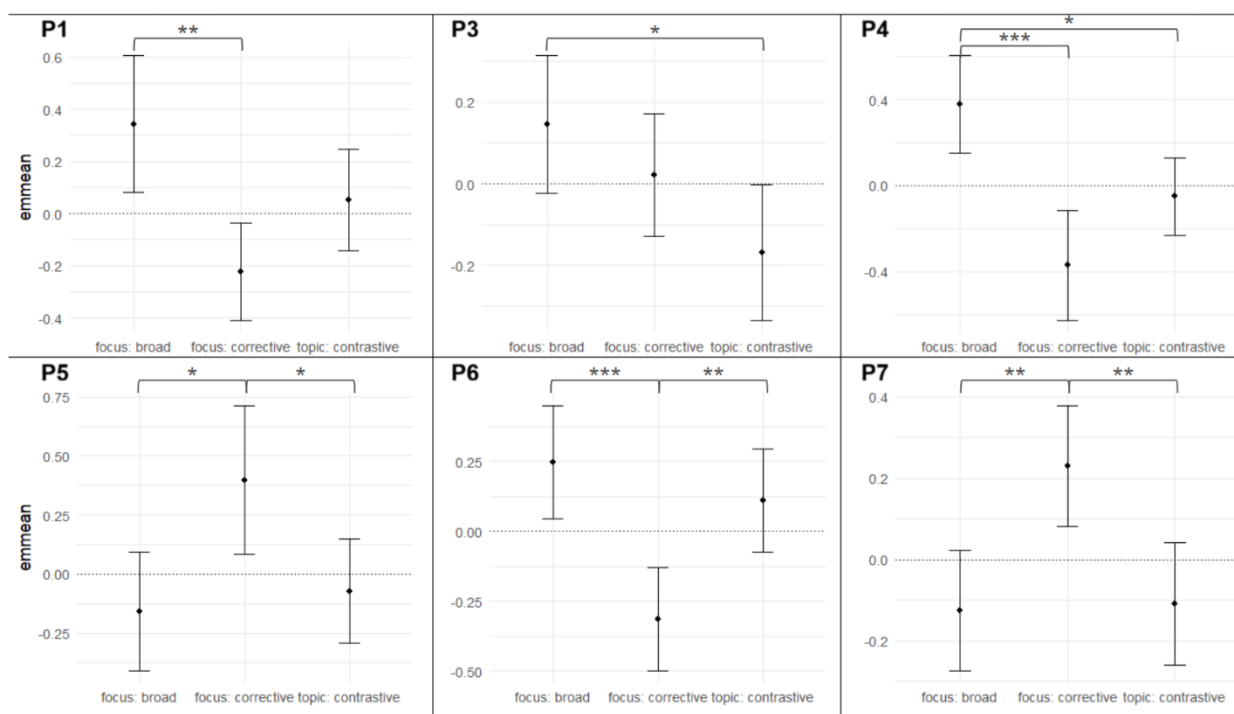


Figure 25. Significant pairwise contrasts between model estimated marginal means of pre-target pause duration across information structure conditions. Significance is indicated by asterisks.

By-participant linear models reveal that information structure type has a significant effect on different gesture phase durations. Pairwise comparisons of model-estimated marginal means for information structure category are significant in the following pointing phase durations (see also **Figure 26** for *p1_point_dur* and **Figure 27** for *p2_point_plat* and *p3_point_re*).

The *p1_point_dur* (pointing onset to target) duration is significantly longer in broad focus than in contrastive topic for participant P3 and P5 (**P3**: $\beta = 1.008$, $SE = 0.326$, $t = 3.095$, $p = 0.009$; **P5**: $\beta = 0.9541$, $SE = 0.375$, $t = 2.546$, $p = 0.04$), and, conversely, it is significantly shorter for participant P7 ($\beta = -0.889$, $SE = 0.282$, $t = -3.149$, $p = 0.007$). For P7, *p1_point_dur* is additionally significantly shorter in broad focus than in corrective focus ($\beta = -1.219$, $SE = 0.282$, $t = -4.320$, $p < 0.001$). The *p1_point_dur* is significantly longer in corrective focus than in contrastive topic for P5 and P6 (**P5**: $\beta = 0.9766$, $SE = 0.352$, $t = 2.772$, $p = 0.02$; **P6**: $\beta = 0.8287$,

$SE = 0.336, t = 2.465, p = 0.05$), and a marginal opposite effect is observed for P4 ($\beta = -0.843, SE = 0.394, t = -2.139, p = 0.09$). No further pointing onset to target duration contrasts were significant.

Information structure effect on pointing duration (onset to target) across participants

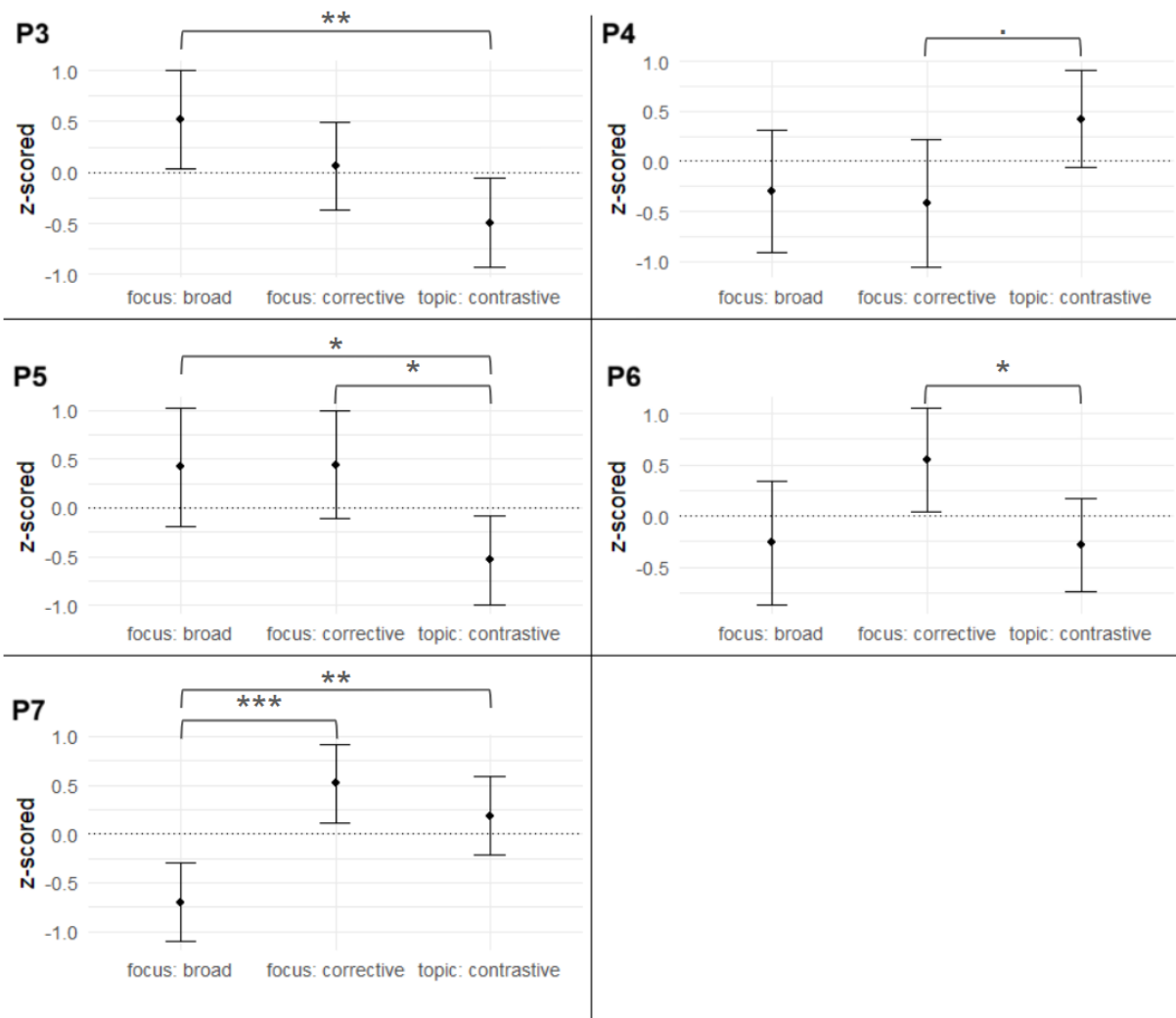


Figure 26. Significant pairwise contrasts between model estimated marginal means of pointing onset to target duration across information structure conditions. Significance is indicated by asterisks.

Information structure effect on pointing plateau and return phases across participants

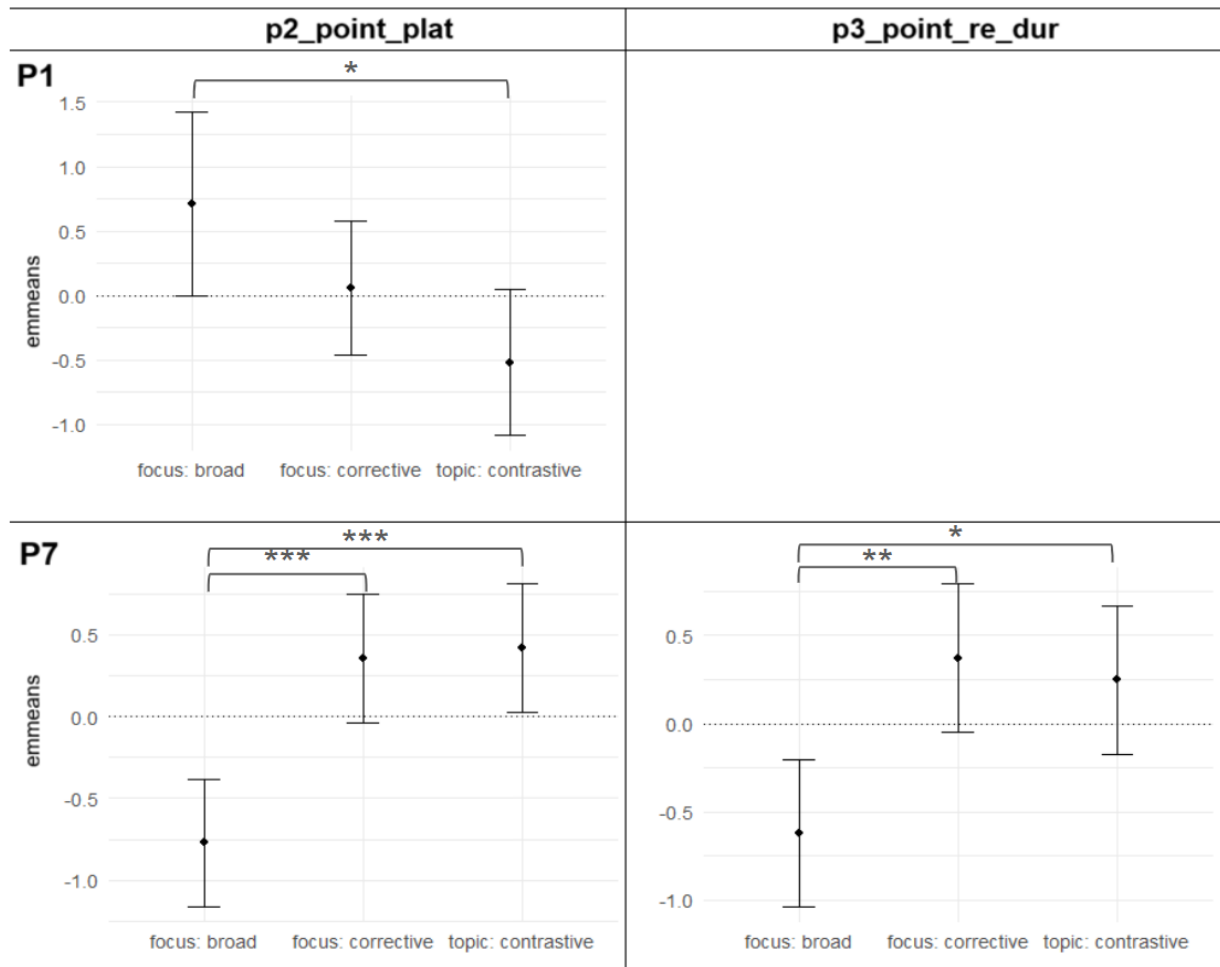


Figure 27. Significant pairwise contrasts between model estimated marginal means of pointing plateau and target to offset durations across information structure conditions. Significance is indicated by asterisks.

The *p2_point_plat* (pointing plateau) duration is significantly longer in broad focus than in contrastive topic for participant P1 ($\beta = 1.229$, $SE = 0.443$, $t = 2.777$, $p = 0.03$). The opposite effect is observed for P7 ($\beta = -1.1909$, $SE = 0.275$, $t = -4.323$, $p < 0.001$) along with a significantly shorter duration in broad focus than in corrective focus ($\beta = -1.1281$, $SE = 0.275$, $t = -4.095$, $p < 0.001$). No further pointing plateau duration contrasts were significant.

Contrasts of *p3_point_re* (pointing return, target to offset) durations are significant for participant P7, with shorter durations in broad focus than in corrective focus ($\beta = -0.992$, $SE = 0.296$, $t = -3.353$, $p = 0.004$), and also in broad focus than in contrastive topic ($\beta = -0.868$, $SE = 0.296$, $t = -2.934$, $p = 0.01$). No further pointing return duration contrasts were significant.

4.3.3 Interaction model (boundary type and information structure) [RQ5]

The interactional effect of initial-boundary and IS condition on speech and pointing phase durations was estimated with by-participant linear models (listed in **Table 13**). Pairwise comparisons were conducted on significant interactions of boundary type and information structure; the contrasts show if the IS contrasts differ across boundary types (interaction effect). **Figure 28** illustrates contrasts for consonant durations, **Figure 29** for vowels, and **Figure 30** for pointing.

Table 13. Linear regression output of significant interactions of IS_SUBTYPE and BOUNDARY on C1, V1, and pointing movement durations.

Movement	Participant	Interaction	Estimate	St.Error	t value	Pr(> t)
		Reference level: broad focus, IP boundary				
<i>a1_c1_forming</i>	P1	IS_subtype focus: corrective boundary w	-1.64	0.7612	-2.154	0.0415
		IS_subtype topic: contrastive boundary w	-1.69	0.7612	-2.159	0.0411
	P4	IS_subtype focus: corrective boundary w	1.43	0.5986	2.394	0.02
<i>a2_c1_plat</i>	P1	IS_subtype focus: corrective boundary w	-3.06	0.8856	-3.452	0.002
		IS_subtype topic: contrastive boundary	-2.26	0.9091	-2.486	0.02
<i>a4_v1_forming</i>	P7	IS_subtype focus: corrective boundary: w	-1.32	0.6213	-2.123	0.04

Movement	Participant	Interaction	Estimate	St.Error	t value	Pr(> t)
		Reference level: broad focus, IP boundary				
<i>a5_v1_plat</i>	P5	IS_subtype focus: corrective boundary w	1.93	0.8531	2.264	0.03
<i>a6_v1_releasing</i>	P4	IS_subtype topic: contrastive boundary w	-1.73	0.7796	-2.213	0.03
	P7	IS_subtype focus: corrective boundary w	1.16	0.5544	2.089	0.04
		IS_subtype topic: contrastive boundary: w	1.43	0.5749	2.494	0.02
<i>p1_point_dur</i>	P5	IS_subtype focus: corrective boundary w	-1.89	0.7819	-2.422	0.02
		IS_subtype topic: contrastive boundary w	-1.53	0.7125	-2.153	0.03
	P6	IS_subtype focus: corrective boundary: w	-2.29	0.6486	-3.525	0.001
<i>p2_point_plat</i>	P6	IS_subtype focus: corrective boundary w	-2.95	0.6201	-4.757	<0.001
<i>p3_point_re</i>	P6	IS_subtype topic: contrastive	-1.499	0.7183	-2.087	0.04

4.3.3.1 Interaction of Boundary and Information structure: C1

The interactional effect of BOUNDARY and IS_SUBTYPE (**Figure 28**) was significant for participants P1 and P4 in the case of *a1_c1_forming* (consonant onset to target) and for P1 in the case of *a2_c1_plat* (consonant plateau). Regarding *a1_c1_forming*, the pairwise contrasts of the interaction effect are not statistically significant for P1 after p-adjustment, and one contrast is marginally significant for P4 ($\beta = 1.433$, $SE = 0.599$, $t = 2.394$, $p = 0.0685$). This latter effect suggests that the boundary effect differs across IS subtypes, with greater durations in the IP boundary than in the word boundary condition for broad focus relative to the difference in corrective focus. Regarding *a2_c1_plat* for P1, one contrast is statistically significant, and one is marginally significant: we find significantly shorter durations in the IP boundary than in the

word boundary condition for broad focus relative to the difference in corrective focus ($\beta = -3.057$, $SE = 0.886$, $t = -3.452$, $p = 0.006$) and also relative to the difference in contrastive topic ($\beta = -2.26$, $SE = 0.909$, $t = -2.486$, $p = 0.06$). No other interactional effects are significant for C1.

4.3.3.2 Interaction of Boundary and Information structure: V1

The interactional effect of BOUNDARY and IS_SUBTYPE (**Figure 29**) was significant for participant P7 in the case of *a4_v1_forming* (vowel onset to target) and for P5 in the case of *a5_v1_plat* (vowel plateau), and P4 and P7 in the case of *a6_v1_releasing* (vowel target to offset). Regarding *a4_v1_forming*, the pairwise contrasts of the interaction effect are not statistically significant for P7 after p-adjustment. Similarly for *a5_v1_plat* for P5, we find no significant contrast after p-adjustment. For *a6_v1_releasing*, the contrasts are not significant in the case of P5, but there is a significant interaction effect in the case of P7, such that the effect of IP boundary relative to word boundary is significantly greater, corresponding to longer durations, in broad focus than in contrastive topic ($\beta = 1.434$, $SE = 0.575$, $t = 2.494$, $p = 0.05$). No other interactions were significant on V1 movement durations.

Boundary and Information structure interaction: C1

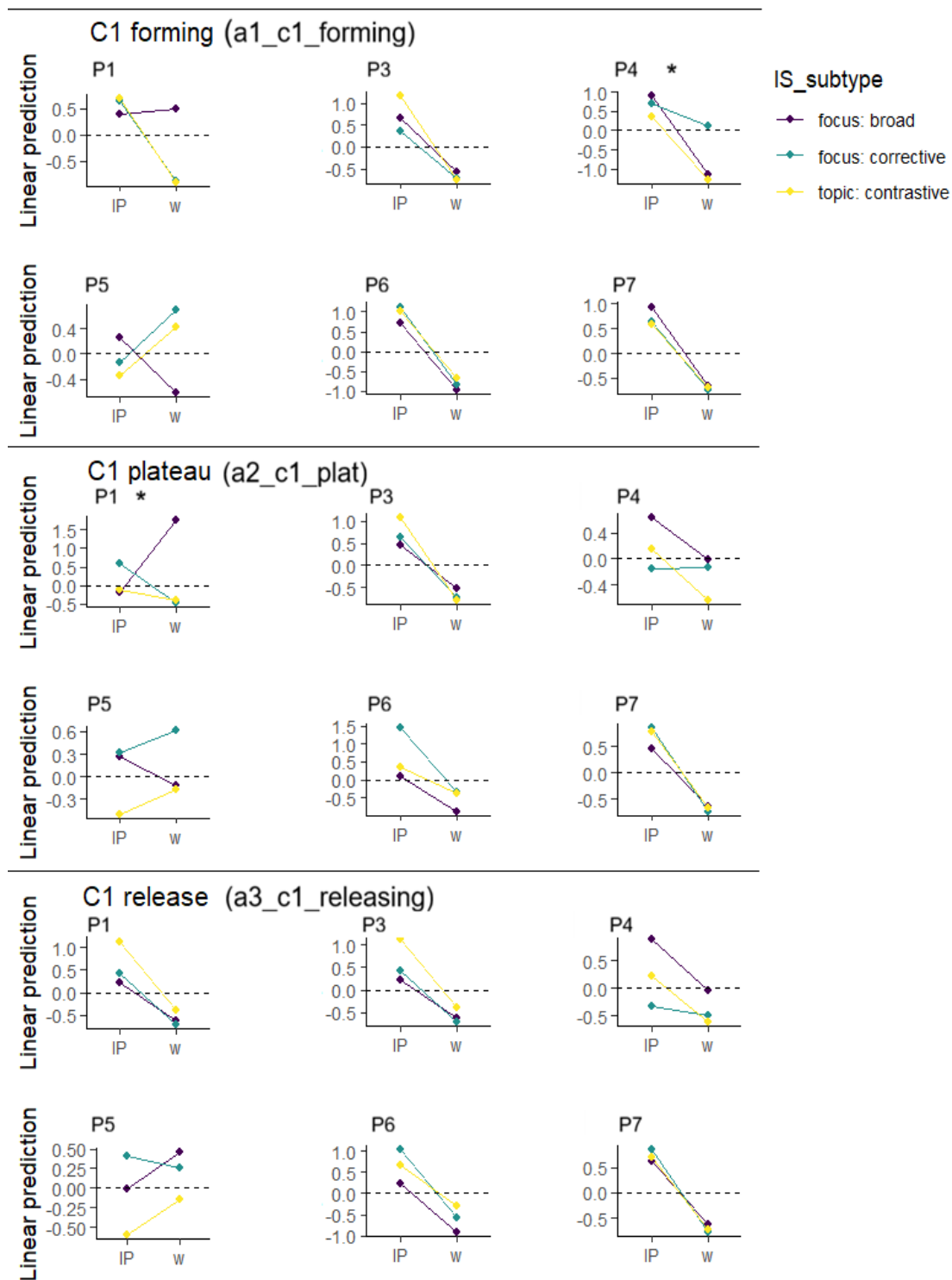


Figure 28. Boundary and IS_subtype interaction effect on C1 movement durations separated by participant.

Boundary and Information structure interaction: V1

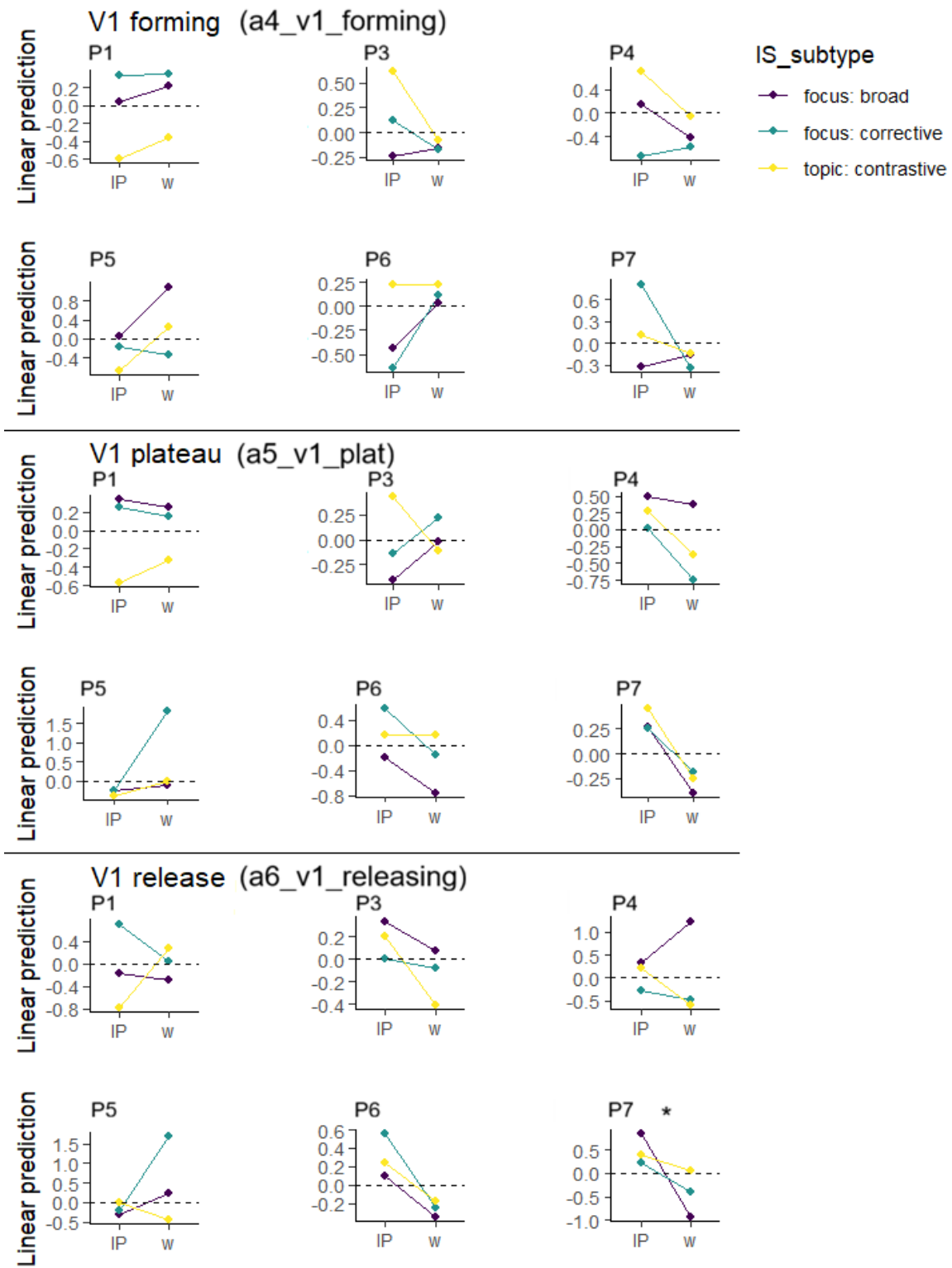


Figure 29. Boundary and IS_subtype interaction effect on V1 movement durations separated by participant.

Boundary and Information structure interaction: Pointing

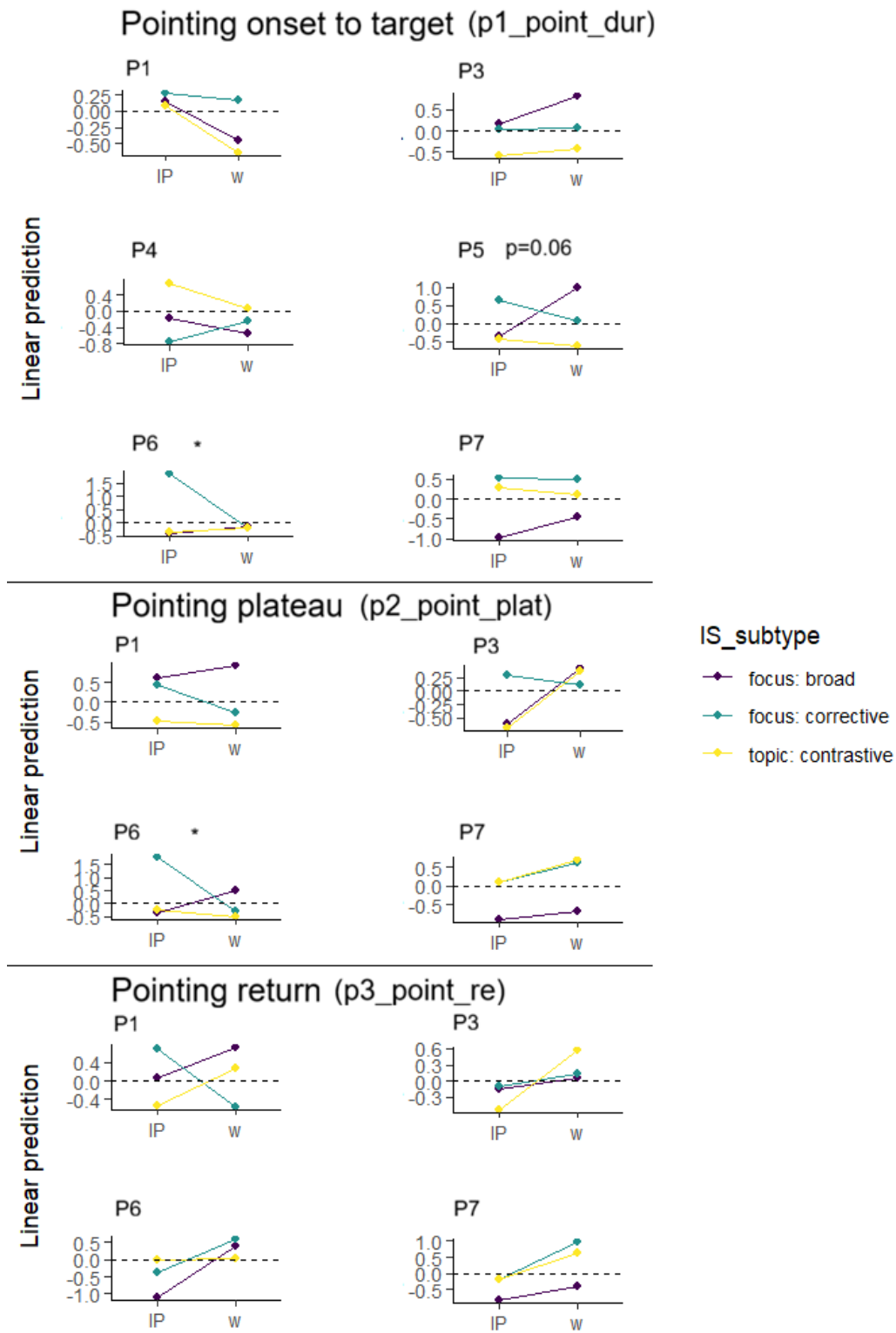


Figure 30. Boundary and IS_subtype interaction effect on pointing movement durations separated by participant.

4.3.3.3 Interaction of Boundary and Information structure: pointing

The interactional effect of BOUNDARY and IS_SUBTYPE (**Figure 30**) was significant for participants P5 and P6 in the case of *p1_point_dur* (pointing onset to target) and for P6 in the cases of *p2_point_plat* (pointing plateau) and *p3_point_re* (pointing target to offset). Regarding *p1_point_dur*, there is a marginal effect in the case of P5 after p-adjustment, with the effect of IP boundary relative to word boundary being marginally smaller, corresponding to shorter durations, in the broad focus condition than in the corrective focus condition ($\beta = -1.89$, $SE = 0.782$, $t = -2.422$, $p = 0.06$). There are significant contrasts in the case of P6: the effect of the IP boundary relative to the word boundary is significantly smaller, resulting in shorter durations, in broad focus than in corrective focus ($\beta = -2.286$, $SE = 0.649$, $t = -3.525$, $p = 0.004$), and the effect of the IP boundary relative to the word boundary is significantly greater, resulting in longer durations, in corrective focus than in contrastive topic ($\beta = 2.151$, $SE = 0.562$, $t = 3.829$, $p = 0.002$). Similarly, regarding *p2_point_plat* for P6, we find two significant contrasts, with a smaller effect of IP boundary relative to word boundary, resulting in shorter durations, in broad focus than in corrective focus ($\beta = 2.95$, $SE = 0.620$, $t = -4.757$, $p < 0.001$), and a greater effect of IP boundary relative to word boundary in corrective focus than in contrastive topic ($\beta = 1.89$, $SE = 0.537$, $t = 3.514$, $p < 0.004$). The *p3_point_re* contrasts are not significant for P6 after p-adjustment, and no other interaction effects are significant for any of the participants.

4.3.4 Post-hoc: Post-target boundary effect

The effect of post-target word boundary (as estimated by the pause) on pointing phase durations was estimated with by-participant linear models. For five of six participants, pause duration is not a significant predictor of the *p1_point_dur* (pointing onset to target) duration, but

in the case of P7, pause durations were inversely associated with pointing duration values ($\beta = -0.01$, $SE = 0.005$, $t = -2.674$, $p = 0.009$). For participants P1 and P3, pause duration is a significant predictor of the *p2_point_plat* (pointing plateau) duration (**P1**: $\beta = 0.004$, $SE = 0.002$, $t = 2.254$, $p = 0.03$; **P3**: $\beta = 0.003$, $SE = 0.001$, $t = 2.629$, $p = 0.01$). For participant P3, pause duration is a significant predictor of the *p3_point_re* (pointing target to offset) duration ($\beta = 0.003$, $SE = 0.001$, $t = 2.617$, $p = 0.01$), and a similar trend is observed for P1 ($\beta = 0.004$, $SE = 0.002$, $t = 1.773$, $p = 0.08$). The effect of the post-target boundary was not estimated for C and V as they are further away from the boundary, and no effect is expected. See **Figure 31** for effects on pointing.

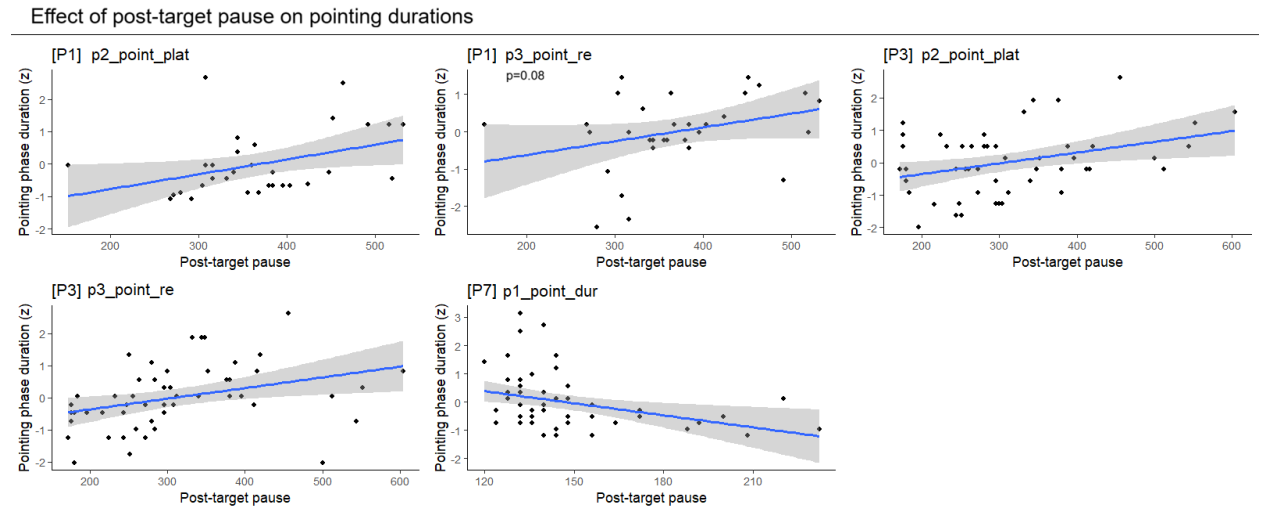


Figure 31. Significant effects of post-target pause on pointing durations. Note, the effect on [P1] p3_point_re is only marginally significant ($p=0.08$).

4.3.5 Post-hoc: Boundary and Information Structure effects on post-target pause

To explore the production of post-target pauses, the effects of initial boundary and information structure conditions were estimated via by-participant linear regression models. There is a significant effect of the initial boundary for all participants, with the word boundary

being associated with longer post-target word pauses in the case of five speakers (**P3**: $\beta = 30.930$, $SE = 9.707$, $t = 3.186$, $p = 0.002$; **P4**: $\beta = 35.737$, $SE = 13.641$, $t = 2.62$, $p = 0.009$; **P5**: $\beta = 103.220$, $SE = 13.340$, $t = 7.737$, $p < 0.001$; **P6**: $\beta = 19.561$, $SE = 4.369$, $t = 4.477$, $p < 0.001$; **P7**: $\beta = 16.506$, $SE = 1.977$, $t = 8.35$, $p < 0.001$), and with shorter pauses in the case one participant (**P1**: $\beta = -32.160$, $SE = 10.016$, $t = -3.211$, $p = 0.001$). **Figure 32.** and **Figure 33** illustrate the differences.

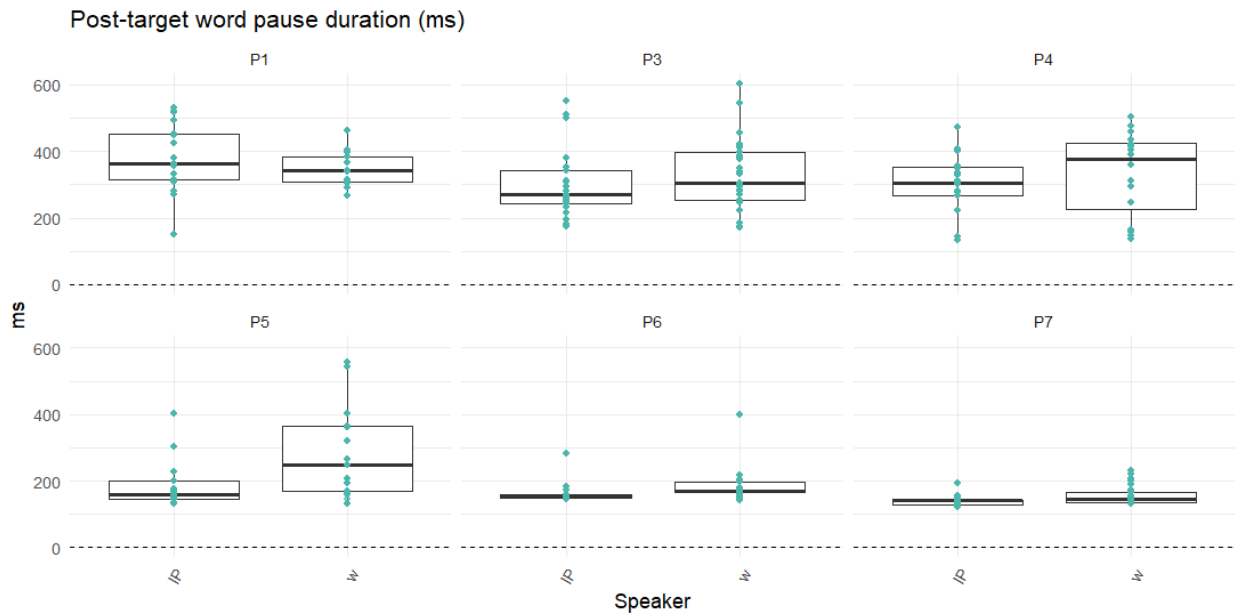


Figure 32. Raw values of post-target pause durations separated by speaker and initial boundary condition.

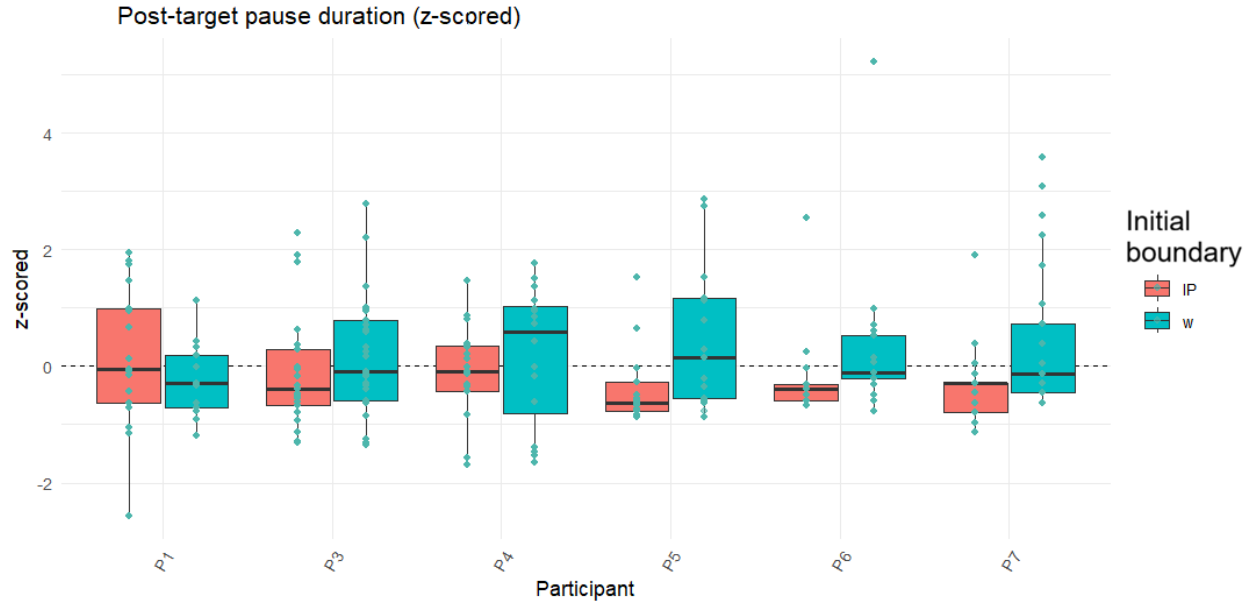


Figure 33. Post-target pause duration separated by participant and pre-target boundary condition.

For all participants, there is a significant effect of information structure condition on the post-target word pause duration as well (see **Figure 34**). Differences were assessed with pairwise comparisons with p-adjustment for multiple comparisons using the Tukey-method (see **Table 14**). For four participants (P1, P4, P5, P7), broad focus is associated with longer post-target pause durations than corrective focus, with a similar trend for P6, and for one participant (P3), the contrast was not significant. For five participants (P1, P3, P4, P5, P7), broad focus is associated with longer post-target pause durations than contrastive topic as well, and the opposite effect is observed for P6. For four speakers (P4, P5, P6, P7), corrective focus is associated with shorter post-target pause durations than contrastive topic, and the opposite effect is observed for P3.

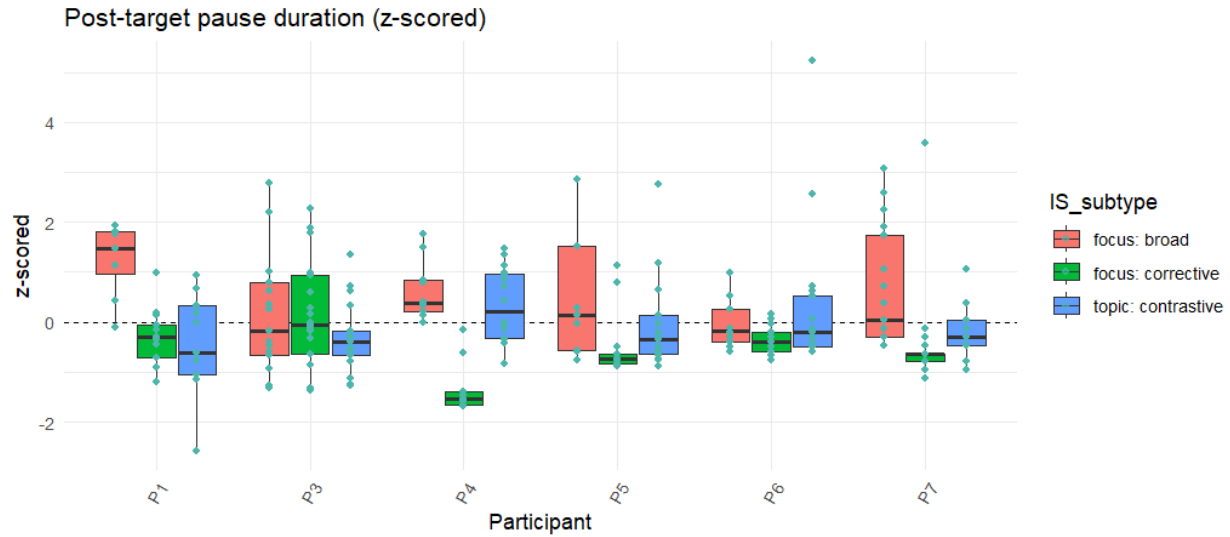


Figure 34. Post-target pause duration separated by participant and information structure condition.

Table 14. Pairwise comparisons of information structure effects on the post-target word pause duration (significant and marginal effects).

	<i>Estimate</i>	<i>Standard error</i>	<i>t-ratio</i>	<i>p-value</i>
<i>broad focus ~ corrective focus</i>				
P1	131.5	9.91	13.273	<0.0001
P4	208.7	11.7	17.784	< 0.0001
P5	109.0	18.5	5.887	< 0.0001
P7	24.84	2.3	10.808	< 0.0001
P6	13.2	5.75	2.295	< 0.06
<i>broad focus ~ contrastive topic</i>				
P1	142.5	10.20	13.973	< 0.0001
P3	45.4	12.2	3.733	< 0.001
P4	31.4	10.3	3.045	< 0.01
P5	63.4	17.4	3.655	< 0.001
P6	-15.0	5.48	-2.742	0.02
P7	18.43	2.32	7.945	<0.001
<i>corrective focus ~ contrastive topic</i>				
P3	44.821	11.5	3.909	< 0.001
P4	-177.4	10.6	-16.665	< 0.0001
P5	-45.6	16.4	-2.787	0.02
P6	-28.2	4.95	-5.702	< 0.0001
P7	-6.42	2.32	-2.766	0.02

Overall, the post-hoc analyses show some evidence that differences in pointing gesture phase durations are associated with differences in post-target pause duration, which in turn are affected by the initial boundary type as well as by the information structure of the target word.

4.4 Discussion

4.4.1 Prosodic boundary effects [RQ3]

The duration of speech and pointing gesture phases were measured following prosodic boundaries of different magnitudes, word and IP. If pointing gestures are affected by prosodic boundaries, durational differences of pointing phases are expected to mirror those of the articulatory gesture phases that the pointing gesture temporally overlaps with. The measured intervals between the onset and target of the pointing gesture overlapped with the target word *Mima*, with a trend for later onset in the IP boundary condition. Accordingly, if the lengthening of the consonant and vowel gestures in the first syllable (*Mima*) is observed, a similar effect would show on these pointing intervals.

Regarding the effects on constriction gestures, results vary. In general, significant lengthening effects of the IP boundary are observed in phases of the C1 gesture (*Mima*) for most speakers (P5 being the only exception, as discussed below). In addition to a general lengthening effect of the IP boundary, in the case of the C1 opening duration (*a1_c1_forming*) an additional interaction with IS_SUBTYPE is observed for P4, with greater lengthening under broad focus than under corrective focus across boundary conditions. Conversely, a shortening effect of IP is observed in the broad focus condition of the consonant plateau (*a2_c1_plat*) for P1. The lack of lengthening found in the production of P5 is likely due to the long pauses the participant produced in the word boundary condition as well (443 ms on average), leading to smaller differences in consonant durations between the boundary conditions.

With respect to the V1 movement durations, no consistent lengthening effect of the IP boundary is observed, with only the release movement (*a6_v1_releasing*, lowering from [i] to [ɐ] in *Mima*) showing significant differences and only for a subset of the speakers. Additionally, in the case of P7, there is a significant interaction effect on the lowering movement, with a greater lengthening effect of the IP boundary in the broad focus condition than in contrastive topic. These results are partially unexpected: while boundary-related consonant lengthening is observed, as expected, the lack of lengthening for vowels was not predicted, as the consonant and vowel onsets align temporally at syllable onsets. Note that under a segmental view of phonology (rather than the gestural view of Articulatory Phonology, or AP), the results could be explained as being related to maintaining a distinction between phonemically long and short vowels. The results however are not readily explainable in the AP framework, as a lengthening effect should be observable at gestures co-extensive with a boundary, regardless of phonemic status.

If pointing gestures are affected by prosodic boundaries, given the results for articulatory gestures, in particular the consonant movement durations, and the position of the pointing gesture, a lengthening effect can be expected in the early pointing gesture phases as well. However, contrary to predictions, no significant lengthening due to boundary is observed in the phase of the pointing gesture that overlaps with the target word (from pointing onset to target) or of the pointing plateau. An exception to this pattern is found in P6's data, where a significant interaction between boundary condition and information structure conditions shows significant lengthening of the pointing duration (*p1_point_dur*) and pointing plateau (*p1_point_plat*) in corrective focus, but not in broad focus or contrastive topic. With respect to the return phase of pointing (from target to onset), which does not generally overlap with the target word, a

shortening effect is observed; in other words, the return phase of pointing is *longer* in the word boundary condition. These results do not support the first hypothesis, i.e. that co-speech gestures lengthen at prosodic boundaries as constriction gestures do.

Three interpretations arise. First, it may be the case that pointing gestures are not under the influence of prosodic structure, which would explain the lack of lengthening near prosodic boundaries. Second, it may be the case that the pointing gesture phases are too far from the boundary to show an effect. Third, beyond interpreting the pattern of results as pointing durations not being influenced by prosodic boundaries, or the pointing gesture being outside the scope of the boundary, however, a potential reason for the lack of lengthening on the pointing gesture is the confluence of the position of the gesture in the different boundary conditions, the shortness of the pre-target word (*ma*), and that word being preceded by an IP boundary (see 4).

[4] IP vs word-boundary condition

context	pre-target	target
(a) [Question] IP boundary	[<i>Ma?</i>]	IP boundary [<i>Mima...</i>]
(b) [Question] IP boundary	[<i>Ma</i>]	word boundary [<i>Mima...</i>]

As mentioned above, an interesting finding is that the pointing gesture onset occurs earlier on average in the word boundary condition than in the IP boundary condition. Thus the IP boundary *preceding* the pre-target word in the word condition may have a lengthening effect on pointing in this condition. Should this be the case, a lengthening effect of pointing in the (planned) IP boundary condition may not be significantly different from the lengthening effect of the pre-target IP boundary, which might then mask any effect of the planned IP boundary.

The unexpected shorter durations of the pointing plateau and return movements for participants P3, P6, and P7 in the IP boundary condition can be reframed as a lengthening effect

in the word-boundary condition, derived from the effect of a *post-target* boundary. These pointing gesture phases were calculated for P1, P3, P6, and P7 (the data was unavailable for P4 and P5 due to labeling differences, as explained in the methodology section). As the post-hoc analysis showed, in addition to the *pre-target* boundary (*ma_Mima*), participants produced pauses after the target word as well (*Mima_be-*). Participants P3, P6 and P7 produced longer *post-target* pauses in the *pre-target* word-boundary than in the *pre-target* IP-boundary condition, and P3 in particular produced phrase boundaries in this position. As theory predicts, the inserted boundary may have a lengthening effect on gestures co-extensive with it, i.e. on the pointing plateau and return movements for participant P3 (and potentially P6 and P7) in the word boundary condition. The *post-target* boundary effects are best illustrated with the pattern observed for P3: the pointing plateau and the return overlap with the boundary, and as shown in the post-hoc analysis, the pointing phases are significantly affected by the *post-target* pause duration. Consistent with this, as the significant main effect of initial boundary type indicated, the pointing plateau is significantly longer, and the return phase is marginally longer in the word boundary condition for this speaker. Similar effects can be observed for P6 and P7 as well (longer return durations and *post-target* pauses in the word boundary condition), although both participants produced shorter *post-target* pauses in general.

A different pattern was observed for P1, who produced shorter *post-target* pauses in the *pre-target* word-boundary than in the *pre-target* IP-boundary condition (see **Figure 32**). While the pointing plateau and return phases are significantly affected by the post-target pause for P1 as well, there was no significant effect of the initial boundary on the pointing movements. Instead, an effect of information structure was observed, with significantly longer pointing plateau

durations in broad focus (**Figure 27**), which corresponded to longer *post-target* pause durations in broad focus for this speaker (**Figure 34**).

Although unplanned, with the gesture phase durations being proportional to the *post-target* pause duration (which varies for one reason or another), these results show a boundary-related lengthening effect on pointing gestures. If pointing gestures were in fact *not* under the influence of prosodic structure, we would not expect to see lengthening due to any prosodic boundary (pre or post target). The overall pattern of results thus indicates that co-speech gestures *can* be affected by prosodic boundaries, potentially lending support to the interpretations above of (a) the pointing phases being outside the scope of the boundary effect or (b) *masked* initial-boundary rather than *no* initial-boundary effects.

4.4.2 Information structure effects [RQ4]

Our second hypothesis concerned prominence-related lengthening of co-speech gestures, again in relation to prominence effects on constriction gestures. If the temporal properties of both types of gestures (speech and co-speech) are affected by prominence similarly, that indicates that prominence structure influences gesture behavior. Vowel and pointing gesture phase durations were examined in broad focus, corrective focus, and contrastive topic conditions for an effect of pragmatic prominence, as greater prominence may correspond to the lengthening of gestures. Additionally, the effect of information structure on pre-target pause duration was examined as well, as longer pause durations may correspond to greater prominence of the upcoming target word (see Mády et al., 2016). Due to their complexity, the results are discussed separately for each participant.

In the case of four of six speakers, no significant effect of information structure was found on the vowel gesture phase durations. Only in P4's production was lengthening observed,

with longer vowel plateau and release movements in broad focus than in corrective focus. In this speaker's case, the *pre-target* pause was longer preceding broad focus than preceding corrective focus (contrary to expectations), and it was longer preceding broad focus than contrastive topic as well. No lengthening of the pointing gesture phases was observed in this direction for this participant, but linear regression indicated a trend for longer pointing durations (onset to target) in contrastive topic than in corrective focus. This lengthening effect may be accounted for by the fact that the participant produced later gestures (**Figure 18**), with the pointing onset to target movement overlapping with the *post-target* juncture, which post-hoc analyses indicated to be significantly shorter in corrective focus than in contrastive topic in P4's production. The lengthening may thus be due to the *post-target* boundary, the magnitude of which is in turn modulated by information structure.

Similarly to P4, a longer *pre-target* pause was observed in broad focus relative to corrective focus in P1's production as well, but pointing phases did not show significant lengthening in this direction. Some information structure effect is shown however on the pointing plateau phase, which was longer in broad focus than in contrastive topic. This observation is consistent with the post-hoc analysis which shows that P1 produced longer *post-target* pauses in broad focus than in contrastive topic, and the pointing plateau overlaps with this juncture. As suggested in the case of speaker P4, the lengthening effect observed for P1 may be due to the *post-target* boundary as well, which is in part modulated by information structure.

In the case of P3, *pre-target* pauses were found to be significantly longer in broad focus than in contrastive topic, and this effect was mirrored in the duration of the pointing gesture's onset to target phase as well. Although no further contrasts were significant for this participant, these results indicate consistency in the effect of information structure.

In P5's production, *pre-target* pauses were significantly longer in corrective focus than in contrastive topic and in broad focus. The pointing duration from onset to target showed a similar effect with longer durations in corrective focus than in contrastive topic. Additionally, the pointing onset to target phase was longer in broad focus than in contrastive topic as well, an effect which may be driven by the *post-target* boundary, which in the post-hoc analysis was shown to be significantly longer in broad focus than in contrastive topic.

In the case of P6, the *pre-target* pauses are significantly shorter in corrective focus than in broad focus and contrastive topic. The pointing onset to target duration does not show lengthening in this direction, but an interaction of boundary type and information structure was found for both pointing onset to target (*p1_point_dur*) and plateau (*p2_point_plat*) durations, with significantly greater lengthening effect of the IP boundary in corrective focus than in broad focus and contrastive topic. For P6, then, a complex interplay of boundaries and information structure is observed in their joint effect on pointing durations, but not in the direction as would be predicted based on *pre-* and *post-target* boundary strengths. That P6 produced longer pointing gestures in corrective focus may be an independent effect of information structure.

In the case of P7, the *pre-target* pause is significantly longer in corrective focus than in broad focus and contrastive topic. The effect of IS on the pause duration is mirrored in the corrective focus vs. broad focus comparison with longer pointing durations (onset to target, plateau, and return) in the former. No significant effect was found on pointing durations in the corrective focus vs. contrastive topic comparison. An additional effect was found, however, with significantly longer pointing durations (onset to target, plateau, and return) in contrastive topic than in broad focus. (Interestingly, despite these differences in pause duration, as the significant interaction of boundary and information structure conditions indicate, the IP boundary effect is

greater in broad focus compared to contrastive topic on the lowering movement of V1
(*a6_v1_releasing.*))

Considering participants together, the hypothesis that the pointing gesture *lengthens* under prominence similarly to constriction gestures is not supported. This is not due to a lack of lengthening of the pointing gestures, however, rather, it is due to a lack of lengthening of the constriction gestures. Effects of information structure on pointing durations were observed, although participants had distinct strategies. The planned comparisons and post-hoc analyses overall reveal the general pattern that the temporal properties of pointing gesture phases are affected by information structure (unlike the stability of coordination with the vowel target, as shown in the previous chapter). The lack of lengthening of consonant and vowel durations indicates that the effect on pointing is not due to the presence of a temporal modulation μ -gesture; rather, the effects may be driven by the insertion of prosodic boundaries, i.e. a π -gesture, the magnitude of which is in turn modulated in part by information structure. Note, however, that P6 presents an interesting challenge to this general pattern.

4.4.3 Interaction effects [RQ5]

The potential interactions of boundary type (IP vs. word) and information structure effects on speech and co-speech gesture phase durations were examined. The interaction effects were discussed in 4.4.1 and 4.4.2 but are summarized here as well. The primary interest in looking at interactions was to probe whether the lengthening effect of prosodic boundaries are greater under greater pragmatic prominence; i.e. whether boundaries and prominence have an additive lengthening effect. To preface the discussion, it should also be noted that the *pre-target* pauses, while modulated by information structure, were not uniformly longer in the expected

direction; i.e., when the target word was pragmatically more prominent (corrective focus, contrastive topic > broad focus).

Overall, while there are some significant interactions between boundary and pragmatic prominence effects, they are not uniform across participants and the interactional effect of pragmatic prominence and boundary magnitude on the C1, V1, and pointing durations does not appear to be uniformly additive. In this light, the marginal interaction effects on the *a1_c1_forming* consonant movement duration for P4 is additive, with the participant producing longer *pre-target* pauses in broad focus than in corrective focus, and accordingly, longer *a1_c1_forming* durations in the *IP boundary – broad focus condition* than in the *IP boundary – corrective focus* condition. In contrast to this, in P1's case (who also produced longer *pre-target* pauses in broad focus), the interaction effect on *a1_c1_plateau* for P1 is not additive in the direction that could be expected from pause duration. Similarly, a lack of additive effect was observed in the case P7, who produced significantly longer pauses in corrective focus than in broad focus and contrastive topic, but the interaction effect on the vowel movement *a6_v1_releasing* did not mirror this pattern. Finally, in the case of pointing phase durations, an interaction effect of pragmatic prominence and boundary was observed for *p1_point_dur* (P6) and *p2_point_plat* (also P6), but not in a way that could be derived from pause durations as modulated by information structure.

These interactions indicate *some* additional effect of pragmatic prominence on boundary related lengthening in some, but not in most cases. These results do not support the hypothesis that boundary and prominence effects on the temporal properties of pointing gestures are additive (although, again, prominence effects were lacking, making an additive effect rather improbable).

4.5 Conclusion

The overall pattern, considering the results of both planned and post-hoc analyses, is consistent with prior findings about boundary- and prominence-related effects on co-speech gestures. While the effect of the boundary *preceding* the target is uncertain (although two proposals were given, both rooted in the scope of lengthening), a lengthening effect of the *post-target* pause is generally observed on the pointing gesture phases that overlap with the juncture; similar effects have been reported in Krivokapić et al. (2017) for English. The effect of pragmatic prominence appears to be more indirect, with information structure affecting pause durations surrounding the target word, which in turn affect co-speech gestures durations (as planned and post-hoc analyses show).

That pointing gestures not only coordinate with speech (as explored in the previous chapter) but are affected by prosodic structure as well suggests a close relationship between the two, underscored by the comparable pattern of results across typologically distinct languages. Additionally, the study provides new evidence for the interplay of prosodic and pragmatic structure in the realization of co-speech gestures, as evidenced by the effect of information structure on boundary strength, which in turn affects pointing gestures. The results thus suggest a strong relationship between co-speech gestures and linguistic structures.

Chapter 5 General Discussion

A substantial body of influential research has established that co-speech gestures temporally align or co-occur with speech in specific ways. Studies have demonstrated that prosodically and pragmatically dominant units of speech anchor co-speech gestures (see e.g. Yasinnik et al., 2004; Rochet-Capellan et al., 2008; Roustan & Dohen, 2010; Ebert et al., 2011; Mendoza-Denton & Jannedy, 2011; Leonard & Cummins, 2011; Loehr, 2012; Debreslioska et al., 2013; Fuchs & Reichel, 2016; Ambrazaitis & House, 2016, 2017; Shattuck-Hufnagel & Ren, 2018; Fung & Mok, 2018; Türk, 2020; Rohrer, 2022; Lehečková et al., 2022; Kügler & Gregori, 2023; Türk & Calhoun, 2023; Türk & Calhoun 2024, among others). In addition, the relative pragmatic prominence of the lexical affiliate has been shown to affect the degree of alignment between the speech and co-speech landmarks (Gregori & Kügler, 2026; Kügler & Gregori, 2023; and suggested in Myslyvyi, Lorenzen, & Baumann, 2026). It has been shown as well that prosodic structure can have an organizing role in gesturing behavior (see e.g. Ferré, 2010; Loehr, 2012; Esteve-Gibert & Prieto, 2013; Esteve-Gibert et al., 2017; Shattuck-Hufnagel & Ren, 2018; Rohrer et al., 2019; Türk & Calhoun, 2023; Rohrer et al., 2023; Türk & Calhoun, 2024) and has an effect on the temporal properties of gestures (Krivokapić et al., 2017).

As this evidence base has grown, it has also highlighted several unresolved questions regarding the precise nature of this alignment. One possible factor contributing to this may be that research has mainly focused on a handful of typologically closely related languages. Notable exceptions, such as Fung & Mok (2018), Rohrer, Hong, & Bosker (2024), and Franich,

Keupdjio, & Nwosu (2025) provide information that shed further light on the role of prosody in alignment, but crucially, they document differences that may be attributed to prosodic typology. With the hypothesis in mind that speech and co-speech gesture alignment is driven by prosodic structure (see Prieto, Esteve-Gibert, and Shattuck-Hufnagel (2025) for a recent review on this), beyond the goal of inclusion for descriptive purposes is the *necessity* of inclusion. If prosodic structure is indeed the driving factor, the lack of prosodic diversity necessarily limits the scope of our understanding, at least under the assumption that languages can, in fact, differ in speech and co-speech gesture behavior (in which case the differences need to be documented, and the patterns of differences need to be better understood).

Another factor is due to studies operating with a largely pre-theoretical notion of alignment, namely temporal proximity, which in and of itself does not speak to the nature of coordination between speech and co-speech gestures. See Leonard and Cummins (2011) for a discussion of this matter, who argue that it is the examination of the variability of the temporal lag with which two events (speech and co-speech gestures in the present case) occur from each other that can elucidate the nature of their co-production. Under this concept of coordination, the hallmark indicating that a co-speech gestural event is aligned with a speech gestural event is the stability of their temporal relationship: the more stable the lag between speech and gesture is, the more likely they are to be coordinated with each other.

With the goal of contributing cross-linguistic evidence to the existing literature on speech and co-speech gesture alignment, **Chapter 3** examined coordination, conceptualized as stability, in Hungarian. In particular, the chapter's central aim was to further our understanding of the temporal relationship between prosodic structure and co-speech gestures (with a focus on pointing gestures), looking into the interplay of prominence and gesturing behavior. The main

research questions the chapter explored were (1) what aspects of speech coordinate with what aspects of pointing gestures in Hungarian, and (2) whether coordination varies significantly as a function of pragmatic prominence, gauged via the inclusion of different information structural categories. Given that Hungarian is a language in which the primary expression of information structure is via word-order modulation, the questions arose as to which speech landmarks co-speech gestures may be most stably coordinated with, and whether coordination changes given additional (and variable) prosodic marking of prominence. Further, the chapter asked whether the *nature* or *source* of prosodic prominence matters in coordination; that is, might coordination vary as a function of information structural category, such as prominence due to the presence of focus vs. topic, or focus and topic types with different relative prominences? Related still to prosodic differences due to information structure, also included in the study was an exploration into the prosodic contours produced on a target word to ascertain if the shape of the contour, potentially linked to information structure, influences gestural coordination.

Next, with the goal of furthering our understanding of the role prosodic structure plays in co-speech gesture behavior, **Chapter 4** examined the role of prosodic boundaries and pragmatic prominence on the temporal realization of pointing gestures. In particular, the chapter asked how prosodic boundaries affect pointing gestures, hypothesizing that if the prosodic structure affects co-speech gestures as it does speech gestures, pointing gestures will exhibit boundary-related lengthening. The chapter also asked if prominence affects gestures in a similar manner (as it does in English, see Krivokapić et al., 2017; Parrell et al, 2014), and examined different types of pragmatic prominence, hypothesizing that if speech gestures lengthen under prominence, pointing gestures will lengthen also, and the degree of lengthening will correspond to the lexical affiliate's degree of prominence. Finally, the chapter asked if prosodic boundaries and

prominence jointly lengthen pointing gestures; that is, if their effects are additive, hypothesizing that the magnitude of the prosodic boundary and the magnitude of prosodic prominence will correspond to the amount of lengthening observed on the pointing gesture. If pointing gestures exhibit such effects, that provides evidence that prosodic structure influences the production of co-speech gestures.

This discussion chapter will discuss the results of the two studies, followed by their interpretation within the framework of Articulatory Phonology, and a discussion of their compatibility with the Kita & Özyürek (2003) model of the co-production of speech and co-speech gestures. The discussion concludes with an overview of the limitations and possible future avenues for research.

5.1 The results: an overview

5.1.1 Coordination

Chapter 3 examined which landmark of speech coordinates with which landmark of co-speech gestures in Hungarian. To approach the question, kinematic, visual, and acoustic data from six speakers was collected. Speakers produced question-answer pairs, where the answer was always the same target sentence containing a target word (*Mima*). Speakers were asked to point to an image of the target word (situated in front of the participant) as they mentioned the target. To examine prominence effects, the questions elicited different information structures (IS) on the target word. These differed both in type of prominence (topic vs focus, narrow and syntactically marked vs. broad and syntactically unmarked focus) and relative prominence within type (contrastive vs. aboutness topic; broad, confirmation, and corrective focus). To examine prosodic prominence effects, f0 contour clustering was performed, where three intonational

contours emerged ((1) high initial peak followed by a fall, (2) low initial f0 followed by a rise, and (3) low initial f0 followed by an earlier rise). The interactional effect of information structure and intonational contour on coordination was examined as well.

The temporal intervals between a number of speech and pointing gesture landmarks were calculated, and the variabilities of the intervals were compared. The interval with the smallest variability is considered the most stable. In selecting the pairs of landmarks, focus was on onsets and targets of both speech and pointing landmarks, and the intervals were all either *onset-to-onset* (i.e. the onset of a speech landmark and the onset of pointing) or *target-to-target* (i.e. the target of a speech landmark and the target of pointing). The speech landmarks selected were constriction gestures (from <pre-target **M**a target **M**ima>): C0 = the onset and target of the pre-target word onset, C1 = the onset and target of the target-word onset, V1 = the onset and target of the first vowel of the target word) and prosodic (pitch accent onset and peak, f0 peak and onset, intensity peak). Of these intervals, **Chapter 3** identified the vowel target of the first syllable as stably coordinating with the target of the pointing gesture (TARGET2__V1_FING), indicating that a target-to-target coordination pattern characterizes the temporal alignment of speech and co-speech gestures in Hungarian.

Regarding the role of information structure, no significant difference was found in coordination. No evidence was found for either the *type* of prominence or for the differences in *relative* pragmatic prominence within category types. That is, coordination is not different under focus vs. background (which we see in comparing coordination under focus vs. topic) or under narrow and syntactically marked vs. broad and syntactically unmarked focus. Relative prominence within information structure categories did not modulate coordination either (contrastive vs. aboutness topic, broad vs. confirmation vs. corrective focus). These negative

results may be interpreted as further strengthening the coordination pattern identified: a lack of variability along information structural lines indicates that the coordination of the pointing gesture with the vowel target is of equal strength regardless of pragmatic differences. Another possible interpretation¹ is that due to the similarity of the communicative context of the experimental task, the communicative need to highlight differentially in the information structure conditions did not arise, as it might otherwise do in more naturalistic communication.

Regarding the role of intonation, no significant differences were found related to the f0 contour type produced on the target word. The study also considered the interaction of intonation and information structure as a potential source of coordination modulation, looking to establish if more prosodic-pragmatic prominence² corresponds to more stable coordination. The results, however, indicate no significant difference along this interactional dimension.

Overall, the stability of the *vowel target* – *pointing target* interval is not modulated either by pragmatic or prosodic prominence or by the interaction of the two, and so it appears to be a rather stable coordination.

¹ Thank you to Marc Swerts and Jonathan Barnes for pointing this out to me (at DGfS26 and Speech Prosody 2026, respectively). See also Zou & Arnhold (2026).

² One reason contour types may differ in their associated prominence is their likelihood of occurring with more or less pragmatically prominent units of speech. Existing research indicates a certain pattern in contour-IS co-occurrence, such as given topics (1) being produced with a rising contour (Genzel et al., 2015), (2) having late peaks when the topic is a proper name (Gyuris & Mády, 2014), (3) producing a plateau on the unstressed syllable (called *tail*) for both aboutness and contrastive topics (but if a rise is produced on the tail, a contrastive topic is more likely, and if a fall is produced, an aboutness topic is more likely (Gyuris & Mády, 2014)), and finally, (4) frequently producing focus with a falling contour (Mády, 2015; Genzel et al., 2015).

5.1.2 Temporal properties of pointing gestures

Chapter 4 examined the temporal properties of pointing gestures in relation to prosodic boundaries and pragmatic prominence. In both cases, the hypothesis was that if pointing gestures are influenced by prosodic structure, as speech gestures are, they will lengthen near prosodic boundaries and under prominence, and the lengthening effect of boundaries and prominence will be cumulative, as is the case for speech gestures. The hypotheses were partially borne out.

5.1.2.1 Prosodic boundaries

To address the question related to the effect of boundaries, the boundary preceding the target word was targeted with the modulation of the question-answer pairs. Participants either produced a word boundary (*Ma # Mima...*) or a prosodic boundary, in particular, an Intonational Phrase boundary (*Ma? Mima...*). As in the previous study, participants pointed towards an image of *Mima* while mentioning the character. The boundary type participants produced was estimated by the length of the pause measured between the pre-target and target words. The pauses indicated that most participants produced the targeted boundaries (and a participant who produced considerably long pauses in the word-boundary condition also kept a distinction between the word and the IP boundaries).

According to the hypothesis, if the pointing gesture is under the influence of prosodic structure, it will lengthen near the prosodic boundary (the IP boundary in the present case) relative to its duration near the word boundary, as will the target-initial consonant and the vowel (*Mima*). Three phases of the pointing gesture were examined for lengthening: pointing onset to target, pointing plateau, and pointing target to offset. Phases that overlap with the effect of the prosodic boundary are expected to show lengthening.

The target-initial consonant showed lengthening following the IP boundary for all participants but one (a participant who produced long pauses before the target word in the word boundary condition), but vowels did not lengthen significantly for most participants (some lengthening was observed for two speakers). Based on the lengthening of the consonant, the lengthening of the pointing gesture that overlaps with the boundary effect (pointing onset to target) was expected. This effect was not observed for either participant, however, which does not support the initial hypothesis.

Three potential interpretations of the results were considered. The first one is that pointing gestures are not showing lengthening because they are not affected by prosodic boundaries, which amounts to pointing gestures not being under the influence of prosodic structure. A second explanation is that the pointing gesture phase begins too late relative to the boundary and therefore it does not overlap (or does not overlap significantly) with the area where the lengthening effect of the prosodic boundary is otherwise observable. Under this interpretation, the pointing gesture may exhibit lengthening due to the presence of a boundary in the right circumstances, i.e. when there is significant overlap. The third possibility is that the pointing gesture in the word boundary condition lengthens due to the proximity of the prosodic boundary preceding the one-syllable pre-target word. This would result in a boundary effect in both boundary conditions, thereby showing no difference between the conditions. The second and third interpretations are consistent with the idea that pointing gestures are affected by prosodic structure and so exhibit lengthening when the conditions for it are met.

Post-hoc analyses showed that such lengthening effects on the pointing gesture are indeed found in relation to another boundary that speakers produced; i.e. the boundary *following* the target word (*Mima_be-*). Participants who produced longer pauses, i.e. bigger boundaries after

the target showed a lengthening on the phases of the gesture overlapping with the boundary (pointing plateau and return). That this effect was observed when the target word was preceded by a *word* boundary explains why a shortening effect in the other condition (pre-target IP boundary) was found in these phases of the pointing gesture. In light of this evidence, it appears that pointing gestures do lengthen near prosodic boundaries.

If pointing gestures were not under the influence of prosodic structure, we would not see the lengthening effect of the post-target boundary. For this reason, the two interpretations that remain regarding the lack of lengthening observed in the IP boundary condition are that the onset to target phase of the pointing gesture was outside the scope of the prosodic boundary target initially, or that the effect was masked by a similar lengthening effect in the word boundary condition due to the earlier IP boundary. Subsection 5.2.2 below interprets the findings in the π -gesture model.

5.1.2.2 Prominence

To address the question related to the effect of prominence, the information structure (IS) of the target word was modulated, as in *Chapter 3*. The target word was either corrective focus, contrastive topic, or it was part of a broad focus response. If prominence has a lengthening effect in Hungarian, more prominence results in greater lengthening of the nucleus of the stressed syllable (the first vowel in *Mi^{ma}*). If prosodic structure affects co-speech gestures, similar prominence effects can be expected on the part of the pointing gesture that overlaps with the syllable (pointing onset to target).

The results indicate that the consonant and the vowel of the stressed syllable do not lengthen as a function of pragmatic prominence; that is, their duration does not change with information structure differences. As such, the lengthening of overlapping pointing gesture phase

(onset to target) is not expected either. Contrary to this expectation, however, we do find that information structure modulates some of the pointing gesture phases.

In most cases, durational differences of pointing can be tied to differences of pause duration (our measure of boundary strength) around the target word (*ma__Mima__be-*), which were found to be modulated by information structure. In other words, it appears that in Hungarian, prominence is not necessarily marked by the lengthening of the prominent syllable, but it is marked by the insertion of prosodic boundaries around the speech unit (as also observed in Mády et al., 2016), and these boundaries affect the duration of overlapping pointing gestures. Cases which cannot be interpreted as boundary effects may arise due to an independent effect of information structure; that is, speakers may mark prominence by modulating the pointing duration directly, as opposed to lengthening only being due to prosodic boundary effects.

The interaction of boundaries and information structure was examined as well, with the hypothesis that if both boundaries and prominence have a lengthening effect, greater boundaries with greater prominence will show more lengthening as well. Beyond the production of boundaries to mark prominence, the interaction of the two did not emerge widely in speakers' data. In only a few instances was an interaction of boundary and information structure significant for pointing gesture phases, and the interaction did not indicate an additive effect.

Overall, the hypothesis regarding the effect of prominence on pointing durations was not borne out in the expected manner, but we observed information structure effect in the modulation of the boundaries before and after the lexical affiliate, and these boundaries having a lengthening effect on the pointing gesture, indicating an indirect effect of information structure on pointing durations.

In summary, the hypothesis of prosodic boundaries having a lengthening effect on pointing gestures is partially supported by the pattern of data. What the source of the boundary may be, that is, whether the lengthening effect is due to the targeted difference between word and IP boundaries preceding the target word or due to the effect of information structure being marked by boundaries is interesting, but in their effect these boundaries do not seem different (which is to say, changes in the source of the prosodic boundary – why the boundary is present – do not lead to different predictions regarding their effects). The evidence presented in *Chapter 4* therefore reinforces the notion that co-speech gestures are influenced by prosodic structure to some extent. The next subsection interprets the findings within the specific framework of Articulatory Phonology.

5.2 The results: an Articulatory Phonology perspective

5.2.1 Coordination

In *Chapter 3*, in Articulatory Phonological terms, the dissertation asked what the coupling relationship between speech and co-speech gestures might be in Hungarian, that is, what the planning relationship between them might be (Saltzman & Byrd, 2000), assuming that stability of coordination indicates such a relationship, even given that co-speech gestures are not part of the lexical representation of words, and are thus post-lexical in nature.

Given the tight target-to-target relationship of co-speech gestures with prosodic prominence in previously studied languages, we initially hypothesized that the co-speech gestures may be coupled with a tone gesture (the pitch accent, or f_0) targets primarily, (**Figure 35/A**). The possibility was also brought up that due to the secondary role of pitch accentuation in prominence marking in the language (secondary to word order modulation that is), the co-speech

gesture may coordinate with another gesture within the prominent unit, that potentially being the vowel as the nucleus of the stressed syllable, which – as pointed out earlier – is a relatively stable and predictable feature of Hungarian (**Figure 35/B**). The pattern of data presented in *Chapter 3* (and reviewed above in *subsection 5.1*) points toward the latter option.

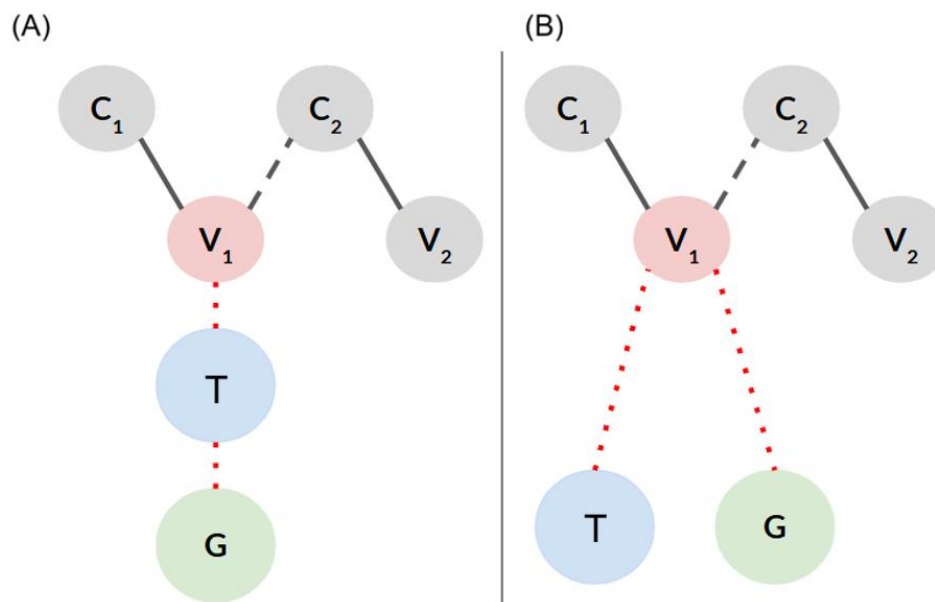


Figure 35. Schematic representation of the co-speech gesture coupling with a tone gesture (A) or with a constriction gesture (B). Notation: C=consonant, V=vowel, T=tonal gesture, G=co-speech gesture.

An alternative reading of the data is that the pointing gesture and the pitch accent (as “abstract linguistic tasks”, (Krivokapić, 2022) coordinate with the prosodic (and also abstract) μ -gesture, which corresponds to lexical stress (Saltzman et al., 2008), as shown in **Figure 36**. This latter view is based on the same observation as the former one, i.e. that both pitch accents and co-speech gestures coordinate with the lexically stressed syllable. Pitch accents are realized on the stressed syllable, and co-speech gestures are either anchored to the pitch accent, which itself is coordinated with the μ -gesture (**Figure 36/A**, as in English), or they are each independently anchored to the μ -gesture (**Figure 36/B**).

As for connecting the μ -gesture with the vowel specifically within the stressed syllable, the assumption in Articulatory Phonology is that it has the greatest temporal effect on the nucleus (for an overview, see Krivokapić, 2022, citing Cho & Keating (2009) who find a lengthening effect of stress on vowel duration in American English and Bombien, Mooshammer, & Hoole (2013) who find the effect of stress in German to be greatest on the nucleus and to decrease away from it).

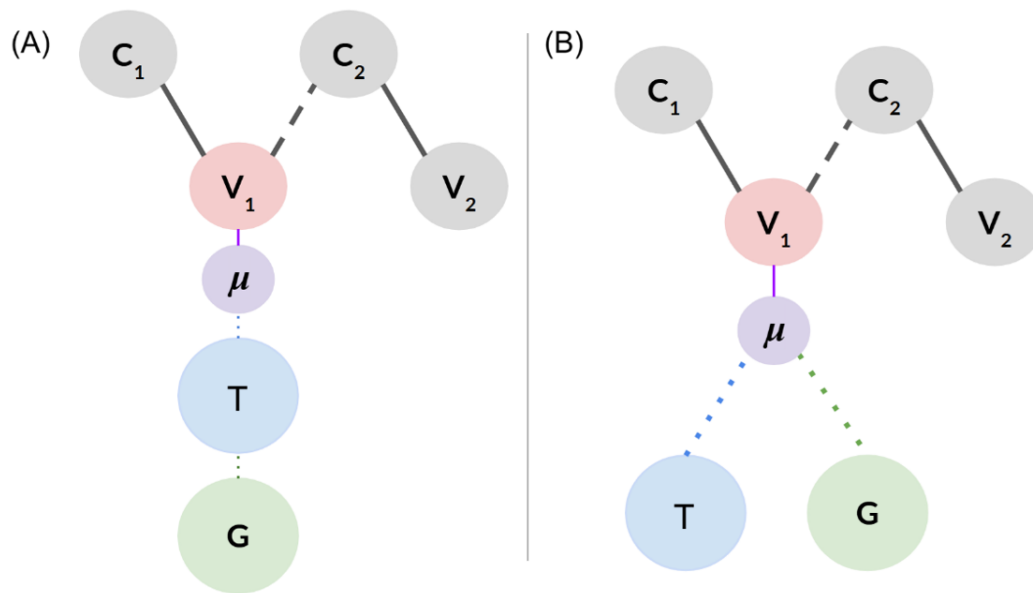


Figure 36. Schematic representation of the co-speech gesture coupled with a tone gesture, which itself is coupled with a μ -gesture (A), and of both the tone gesture and the co-speech gesture coupled with a μ -gesture (B). Notation: C=consonant, V=vowel, T=tonal gesture, G=co-speech gesture, μ =temporal modulation gesture.

The difference between the views depicted in **Figure 35** and **Figure 36** is largely theoretical at present, in that neither explains the coordination results with more adequacy. The former view leaves room for the interpretation of coordination with the vowel being due to the vowel's status as the nucleus of the phrase-initial syllable (which also happens to be stressed). In support of the view is the evidence shown in **Chapter 4** that pragmatic prominence is marked by the insertion of boundaries, which indicates that phrase edges can be associated with

prominence, lending prominence to the phrase-initial syllable. This is rather uncontroversial, but in Hungarian, as mentioned before, the phrase-initial syllable coincides with the stressed syllable, making the motivating factor in coordination difficult to determine.

The latter view, coordination with a μ -gesture, makes the assumption explicit that it is due to lexical stress that the co-speech gesture coordinates with the vowel in Hungarian (which happens to be in the first syllable). Cross-linguistic evidence from languages in which the co-speech gesture coordinates with pitch accents may support this interpretation of the data, as pitch accents are realized on lexically stressed syllables (as is the case in English, for instance). Additionally, as argued in *Chapter 3*, as a predictable feature of the language, lexical stress is not an unlikely anchor.

Whether there is indeed a temporal modulation gesture present in Hungarian is not clear, however, as reports on the phonetic correlates of stress in Hungarian vary, with Markó et al. (2019), Mády & Kleber (2010), and Vogel, Athanasopoulou, & Pincus (2015) not finding lengthening to be a correlate and Szalontai et al. (2016) finding it to be so but attributing the effect to other factors. In languages with stress-based prominence, lengthening can also be expected due to phrasal prominence, see e.g. Mücke & Grice (2014) for German and Krivokapic et al. (2017) for English, and references therein. In Krivokapic et al. (2017) in particular, cumulative lengthening was found corresponding to differences in information structure (in order of increased lengthening: deaccented word < broad focus < narrow focus < contrastive focus), indicating the possibility of different magnitudes of a μ -gesture. As reported for Hungarian in *Chapter 4*, no lengthening of the vowel was found as a function of phrasal prominence, which could indicate the absence of a μ -gesture. A conclusion regarding lengthening (and the presence of μ) due to lexical stress cannot be made for Hungarian, however, as the first syllable was

lexically stressed in all conditions, and so comparison with an unstressed syllable is not available. What can be said is that if a μ -gesture is present, it is not due to phrasal prominence.

It may be the case that instead of a temporal modulation gesture (μ_T), a spatial modulation gesture (μ_S), is present in the lexically stressed and prominent syllable. Saltzman et al. (2008) describe spatial modulation gestures as either strengthening or reducing the movements associated with constriction gestures (and see e.g. Cho et al. (2016) for a boundary-related effect of μ_S in Korean). Examining Hungarian, Markó et al. (2018) and Markó et al. (2019) find reduced formant variability associated with prominence, and Vogel et al. (2015) find a lack of short vowel centralization under focus. Inasmuch as such phenomena can be understood as spatial strengthening under prominence, there may be support for a spatial modulation gesture in Hungarian. The dissertation focused on coordination and the temporal properties of co-speech gestures, but future research could explore this possibility.

5.2.2 Temporal properties

In *Chapter 4*, the dissertation asked if pointing gestures are under the influence of prosodic structure, and more specifically, prosodic boundaries (the π -gesture) and prominence (a μ -gesture). As explained in the previous subsection, the presence of a μ -gesture cannot be inferred from the available data, this subsection will therefore focus on the lengthening effects attributed to the π -gesture.

As explained above, if co-speech gestures are affected by the π -gesture, the lengthening of co-extensive phases of the gesture are expected, and the amount of lengthening is predicted to be proportional to the magnitude of the π -gesture. In *Chapter 4*, this magnitude was inferred from the duration of the pause at the given juncture. As such, there is no categorical distinction

between π -gestures (Byrd & Saltzman, 2003), the stronger the boundary, the greater the magnitude of the π -gesture, and the greater the expected lengthening.

In our data, participants produced prosodic boundaries in the IP boundary condition, maintaining a distinction between that and the word boundary at the targeted juncture preceding the target word. For this reason, the data was analyzed along the lines of having an IP boundary vs. a word boundary, which is a categorical distinction, but the underlying assumption is the presence vs. absence of a π -gesture, respectively. In addition to this, participants produced prosodic boundaries following the target word, and so the data can be interpreted with the presence of an additional (post-target) π -gesture. The duration of the pause from which this boundary was inferred was found to be related to information structure; that is, participants appear to mark information structure with the insertion of a π -gesture, the magnitude of which is speaker-specific but varies along information structural lines.

As reviewed in the previous section, lengthening of the consonant (though not of the vowel) was found related to the boundary preceding the target word, which indicates the presence of a π -gesture. The lack of lengthening of the vowel is unexpected and is not predicted by the model, as the consonant and the vowel begin simultaneously, and so comparable lengthening of both can be expected (a possible explanation is that the vowel is longer than the consonant and is less overlapped with the π -gesture overall). No lengthening of the pointing gesture was found either, however, for which three possible interpretations were given. The first interpretation is that co-speech gestures are not under the influence of prosodic structure, in which case no lengthening is expected. As also reviewed above, lengthening was found related to the π -gesture following the target word, which is incompatible with the interpretation that boundaries do not affect co-speech gestures. For this reason, this first interpretation seems

unlikely. The other two interpretations take into consideration the position of the pointing gesture relative to the π -gesture: either the pointing gesture begins outside of the scope of the π -gesture, or the π -gesture present preceding the pre-target word (i.e. between the context setting question and the first word of the answer, “ma”) lengthens the pointing gesture in the word boundary condition and the π -gesture immediately preceding the target word lengthens the gesture in the IP boundary condition, as illustrated by **Figure 37** below. The latter two views are compatible with the interpretation that prosodic structure affects co-speech gestures.

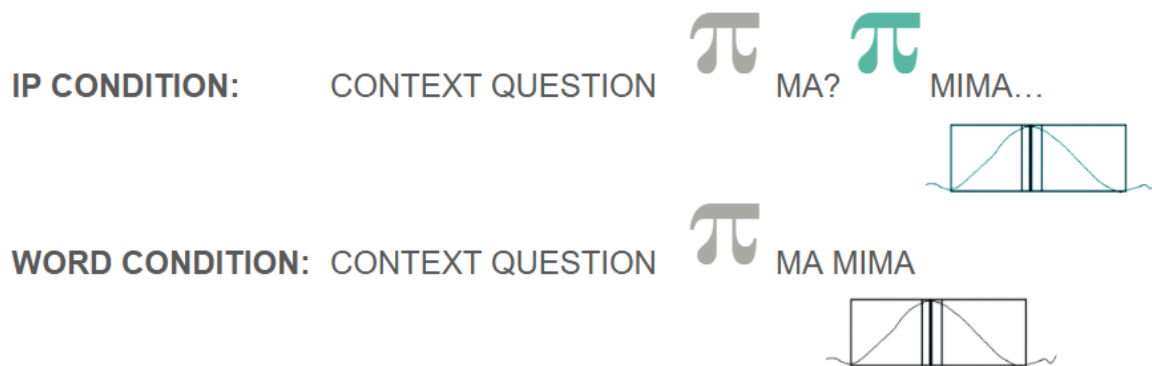


Figure 37. Schematic representation of π -gesture effects in the IP and word boundary conditions.

To summarize, it was shown in **Chapter 3** that pointing gestures coordinate with the vowel in the first, lexically stressed syllable of the lexical affiliate, and coordination was found not to be influenced by information structure. Next, in **Chapter 4** it was shown that pointing gestures lengthen under prosodic boundaries (the π -gesture), and the magnitude of prosodic boundaries was modulated by information structure in speaker-specific ways. Overall, the two studies presented provide evidence that prosodic structure influences co-speech gesture behavior in Hungarian.

5.3 Language production and temporal alignment

It has been well documented in a broad body of work that co-speech gestures and speech are closely connected, and several production models have been proposed to capture aspects of their co-production. Having demonstrated cross-linguistic evidence that prosodic structure affects deictic co-speech gestures, the question continues to present itself as to *how* this temporal alignment arises (both in terms of coordination and in terms of changes in temporal properties) between prosodic structure and co-speech gestures in production. As pointed out by Krivokapić et al. (2024) and Türk & Calhoun (2024), models of co-speech gesture production do not engage with the question of temporal alignment with speech (McNeill, 1992, 2005; de Ruiter 2000; Kita & Özyürek, 2003); for an overview of the models, see Wagner et al. (2014).

The present discussion will focus on the Interface Hypothesis model of Kita & Özyürek (2003) which is based on Levelt (1989)’s speech production model and the Growth Point theory (McNeill 1992, 2005; McNeill & Duncan, 2000). The former specifies the following stages of speech production: conceptualizer (where the message to be communicated is determined), formulator (where the linguistic structure of the message is determined), and articulator (where the execution of the message is developed). The latter argues that both speech and co-speech gestures have a common origin (called the “growth point”), which serves as an explanation for the shared conceptual aspects of the two.

Figure 38 below shows the model proposed in Kita & Özyürek (2003). The figure illustrates the common origin of speech and gesture in the *Communication Planner*, which is where the approximate ordering of the message is proposed to happen, along with the determination of modality. This stage is informed by the *Discourse Model* (highlighted in blue), where the goal of the discourse is decided, and, relatedly, what aspects of the message should be

made more or less prominent. As such, this stage is related to information structure (see Türk (2020) for a discussion). Next, from the *Communication Planner*, parallel stages plan the speech and the co-speech gesture in the *Message Generator* and *Action Generator*, respectively. As the arrows in the figure indicate, these two stages influence each other (see e.g. Mol & Kita (2012) for the *gesture* → *speech* direction and Kita & Özyürek (2003) for the opposite). In the co-speech gesture pipeline, the next step is the planning of the gesture execution at the *Motor Control* stage, followed by the gesture output. For speech, the *Formulator* is next, where linguistic structures (symbolic aspects of language) determine the language output. The *Formulator* itself is informed by the *Lexicon*. Not pictured is a further step presumed in the speech production pipeline between the *Formulator* and the output; namely, an *Articulator* or a *Motor Control* stage for speech that parallels the *Motor Control* stage for co-speech gestures. See the Extended Interface Hypothesis of Türk (2020) for a discussion of this, as well as of an additional stage, the *Action Formulator* between the *Action Generator* and the *Motor Control* stages.

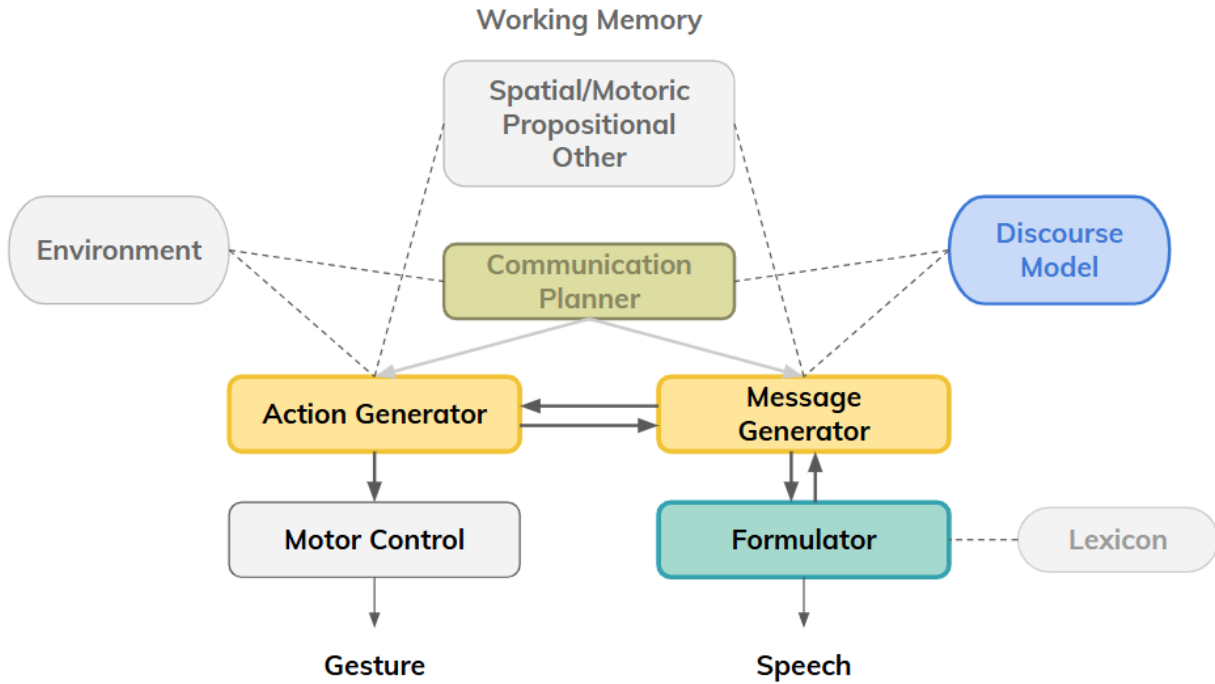


Figure 38. Adapted from Kita & Özyürek (2003) to highlight areas of focus for prosody (in the Formulator) and information structure (in the Discourse Model).

The model is underspecified regarding the temporal relationship (coordination and temporal properties) between speech and co-speech gestures. Importantly, however, Kita & Özyürek (2003) propose a line of feedback from the *Formulator* to the *Action Generator* via the *Message Generator*, which allows for the influence of linguistic structure on co-speech gesture formulation. With prosodic structure being planned in the *Formulator* (based on Levelt, 1989), the alignment phenomena between prosodic structure and co-speech gesture realization could be possible via this pipeline. The Extended Interface Hypothesis of Türk (2020) goes further by proposing a direct line between prosody and gesture. Prosody is planned in particular in a *Grammatical Encoder* of the *Formulator* (building on Calhoun, 2010) with feedback from a *Phonological Encoder* (also in the *Formulator*), and co-speech gestures are planned in the *Action Formulator* (which, as mentioned earlier, takes input from the *Action Generator* and outputs to

the *Motor Control* module). The two are connected via links from the *Formulator* to the *Action Formulator*, through which phrasal and prominence-related synchronization is achieved.

The results of the studies presented in **Chapter 3** and **4** of the dissertation support there being such a link between the *Formulator* and Türk (2020)'s *Action Formulator*. As reported, the studies find effects of prosody on the realization of pointing gestures: their relative timing and temporal properties. Regarding relative timing, as reviewed previously, the coordination between speech and co-speech gestures shows some cross-linguistic differences, likely related to prosodic typology. These differences can be well accommodated within Türk (2020)'s extension of the Interface Hypothesis, as the *Formulator* determines the language specific aspects of prosody, such as prominence marking. Similarly, the lengthening effect of prosodic boundaries on co-speech gestures can be accounted for via the feedback link between the *Formulator* and the *Action Formulator*. Further evidence (as discussed in the introduction) for cross-linguistic differences in temporal alignment between prosodic structure and co-speech gestures is due to Türk (2020) and Türk & Calhoun (2024) for instance, who find co-speech gesture phrases (units of gestures composed of gesture phases, such as onset-stroke-return; see **subsection 1.2** in **Chapter 1** for more information) to be co-extensive with prosodic phrases (groups of ips, or intermediate phrases) in Turkish, and Loehr (2004) who reports multiple gesture-phrases to align with IPs in English. As Türk & Calhoun (2024) argue, such cross-linguistic differences in alignment may be rooted in the language-specific organization of prosodic structure. Cross-linguistic differences can not only be accommodated in the Extended Interface Hypothesis model, but they strengthen the idea of an interaction between the *Formulator* and the *Action Formulator* proposed in Türk (2020).

Research related to further links between speech and co-speech gestures is an active area of inquiry rooted in the biomechanical processes involved in speech and language production. Studies find cross-modal effects that provide evidence for entrainment between the two modalities (see e.g. Rusiewicz et al., 2014; Parrell et al., 2014; Pouw, Harrison, & Dixon, 2020a; Pouw, Harrison, & Dixon, 2020b; Pouw, Harrison, Dixon, 2020c; Pouw, Paxton, Harrison, & Dixon, 2020; Pouw et al., 2021; Pouw & Fuchs, 2022; Pouw et al, 2025, Turk et al, 2026, and works cited within). Findings include rhythmic arm movement entraining with f_0 and the amplitude envelope (Pouw, Harrison, & Dixon, 2020a), arm weight and movement modulating amplitude (Pouw et al, 2025), and head position modulating f_0 (Ćwiek & Fuchs, 2025). While such biomechanical effects may apply universally (cross-linguistically) by virtue of the mechanics of speech production being affected, they do not rule out the presence of a linguistic link between prosody and gesturing. In accordance with the latter type of link, coordination patterns across languages do not appear to be universal, as shown in the observed typological differences (see also Franich et al. (2025) on this point). Recent research in support of this show the influence of native language lexical stress patterns on hand movements in L2 speakers (Bosker et al., 2026) and language-specific speech and co-speech gesture timing patterns in bilingual speakers of languages with different prosodic typologies (Franich, 2026). Which aspects of speech-gesture co-production are due to biomechanical factors, and which ones are due to linguistic factors remain to be disentangled, but the pattern of results across languages indicates the presence of a linguistic link.

Regarding the role of information structure (IS), as reviewed in the introduction, co-speech gesture frequency effects, gestural prominence effects, and temporal patterns have been found, suggesting a connection. For a discussion of how information structure relates to speech

and co-speech gesture production, see the *Extended Interface Hypothesis* model of Türk (2020). The studies in the present dissertation tested IS effects on coordination and on the temporal properties of gestures. As discussed, no IS-related effects on coordination were observed in Hungarian, which suggests that synchronization is more likely to be driven by prosody. These results differ from prior observations of focus modulating the strength of alignment (as reported in Gregori & Kügler (2026) and Kügler & Gregori (2023) for German) and should be explored further with the examination of other typologically distinct languages. A lengthening effect of IS was observed, which was attributed to an IS effect on prosodic phrasing. From the data available therefore, only an indirect link could be established between information structure and gesture timing. This finding is consistent with that of Türk & Calhoun (2024), who observe IS effects on gestures to be mediated through prosodic phrasing in Turkish.

To conclude, although the Interface Hypothesis of Kita & Özyürek (2003) is underspecified for the role prosody and information structure plays in speech and co-speech gesture co-production, it allows for the exchange of information at multiple stages. In agreement with prior studies, and in particular with Türk (2020)'s Extended Interface Hypothesis, the dissertation provides some evidence for a link between the *Formulator* for speech and the *Action Formulator* for gestures, shown in patterns of coordination rooted in prosodic typology and in the effect of prosodic structure on the temporal properties of pointing gestures. As for IS, although the studies on Hungarian only provide evidence for an indirect link between information structure and pointing gesture realization, other work suggests further effects of IS, which can be modeled via the *Discourse Model – Conceptualizer* link (see Türk, 2020). If a model is to account for the prosodic, information structural, and interactional effects, future work will have to further specify how the (direct and indirect) links work together.

5.4 Notes on Hungarian prosody

Although it was not the intention of the dissertation to make claims about the prosodic typology of Hungarian, the data presented in *Chapter 3* and *Chapter 4* may warrant certain observations regarding intonational and prosodic phrasing patterns.

The intonational contours participants produced (as was shown via contour clustering) in the different information structure conditions support prior observations. As a reminder, the examined contours are produced on the entire target word (*Mima*, stressed + unstressed syllable) and these observations do not pertain to the stressed / accented syllable alone. Participants tended to produce a falling contour (high f_0 in the first syllable followed by a fall and flat f_0) in the narrow focus conditions, similarly to the contours reported in Mády (2015) and Genzel et al. (2015). In the topic conditions, participants had a tendency to produce an overall rising contour (initial low f_0 followed by a rise either in the second and final syllable or around the syllable boundary), which is fairly consistent as well with prior observations (see Gyuris & Mády (2014) for given and contrastive proper name topics). In the broad focus condition, participants mainly produced rising contours as well – a reason for this may be that they realized the pre-verbal target word *Mima* as a new topic (thereby creating a topic + predicate structure, which is not uncommon in Hungarian – see the overview of information structure marking in the introduction). Should this be the case, unlike in Genzel et al. (2015), the information status (given/new) of the target word in topic position did not modulate the intonational contour. Overall though, findings corroborate prior observations about general tendencies in the intonational marking of focus and topic in Hungarian.

With respect to prosodic phrasing, the results reported in *Chapter 4* indicate that speakers insert and/or strengthen prosodic boundaries around the target word in a manner that is

modulated by information structure. One interpretation of these observations is that prominence is marked by phrasing in Hungarian, which may have indications for the prosodic typology of the language, placing it among phrasal or Edge prominence languages, such as Korean (see Kügler & Calhoun (2020) and Jun (2005, 2014, 2025) for reviews of prominence marking strategies across languages). A further consequence of this interpretation is that the intonational peaks on the edge of the target word may be classified accordingly as boundary tones (BT) marking the prosodic boundary rather than pitch accents (PA); see also the introduction in Kügler et al. (2026) for a similar point.

The coordination results may provide some insight into the matter as well³ (see Loehr (2012) for instance, where the intermediate phrase level in English is motivated by co-speech gesture alignment patterns). As reported in *Chapter 3*, the pointing gesture coordinates with the nucleus of the stressed first syllable of the target word. It was noted that the landmark is an ambiguous anchor in the sense that, given the design and given the left-headed prosodic structure of the language, it is not clear if the pointing gesture coordinates with the vowel due to the syllable's phrase-initial position or due to it being stressed. Either way, however, coordination with a stable landmark on the left edge of the target word (rather than with tonal targets) provides further support to prior proposals that the boundary has a prominence-lending function in Hungarian, as also suggested by Reichel et al. (2016) and Mády et al. (2016). Whether this pattern necessitates the categorization of the language as one with (only) Edge prominence marking is not certain.

Returning to the boundary insertion/strengthening: consistent also with the results is the view that a function of the inserted boundaries is to mark the division between topic(s) and the

³ Thank you to Frank Kügler for pointing this out to me.

comment. Participants tended to produce larger pauses following the target word in the contrastive topic and broad focus conditions. As mentioned in *Chapter 4*, larger pauses can indicate larger boundaries, which – if speakers indeed interpreted and realized the target word as a new topic in the broad focus condition as well, as suggested by the intonational contour patterns – can correspond to the topic-comment boundary. Following this interpretation, with respect to the pre-target pause, we might expect the greatest boundary to precede corrective focus, as it marks the onset of the comment; this however was not consistent among speakers. Where differences were significant, broad focus and corrective focus were preceded by greater boundaries (with only one speaker producing greater boundaries in contrastive topic than in broad focus, and none producing greater ones in contrastive topic than in corrective focus). One explanation to this pattern may relate to the information status of the target unit (new in broad and corrective focus, but given in contrastive topic), which can modulate prominence (see Genzel et al. (2015) who finds givenness to have an independent effect on the realization of the background in Hungarian). Overall then, the inserted/strengthened boundaries can mark the topic/comment boundary (post-target word: broad focus, contrastive topic > corrective focus), but they may also be modulated by the relative prominence of the target unit (perhaps due in part to differences in information status; pre-target: broad focus, corrective focus > contrastive topic).

This latter interpretation leaves open the possibility that the target-initial peaks are either pitch accents or boundary tones, and the target-final peaks are boundary tones. This reading of the data is consistent with the classification of Hungarian as a Head/Edge prominence language. According to Jun (2005, 2025), such languages mark both the heads and the edges of prosodic units (prosodic words (PW) or accentual phrases (AP)) and typically have fixed stress with weak acoustic correlates. Hungarian appears to fit this description with fixed stress and intonation

marking on the first syllable of APs, and also a BT on the last syllable. Further though, Jun (2025) in particular suggests that one edge is marked by a pitch accent and the other edge by a boundary tone. Adhering to this account, the intonational peak on the first syllable of the AP can be interpreted as a PA and that on the final syllable as a BT.

One tentative synthesization of the two approaches may be that boundaries (and boundary tones, or edge marking) mark the broad information structure of the utterance ([topic]+[comment]), as can be expected in a discourse configurational language with structural prominence marking, and the strength of the boundary may be modulated further by the relative prominence of the topic and comment phrases (see Reichel et al., 2016; Mády et al., 2016). Additional head marking (via pitch accentuation) arises when further emphasis is necessary. Evidence for prominence marking in addition to focus-phrase marking was reported in Langer & Kügler (2023), where speakers produced additional f₀ peaks on the narrow-focused item within a larger pre-verbal focus phrase. These distinct peaks could be interpreted as edge and head markings. (If, however, speakers insert boundaries within such complex NP (or DP) focus phrases immediately preceding the narrow-focused item, it may be a strong indicator of Hungarian falling in the category of languages that mark prominence via phrasing, i.e. Edge prominence languages. An additional question to resolve is what such nested prosodic phrasing may imply for the architecture of the Prosodic Hierarchy.)

Based on the data available we cannot firmly place Hungarian among the Edge prominence languages as the results are consistent with the alternative Head/Edge prominence category as well. While the results reported in *Chapter 4* indicate that Hungarian marks prominence via phrasing, supporting prior observations, the phonological status of the initial peak(s) remains unclear. Whether boundary insertion/strengthening is the only prosodic strategy

available to Hungarian speakers is unclear as well, and future work is necessary to determine if or when additional strategies, such as pitch accentuation, are available (if the initial peaks are interpreted as BTs, that is) and what precisely modulates the combination of the different strategies.

5.5 Limitations and future directions

The studies can be expanded in numerous directions. *Chapter 3* has touched upon some of the limitations of the coordination study, identifying as a future direction the examination of the role of f_0 in gestural coordination in Hungarian in contexts where prosody marks information structure. In particular, the prosodic marking of narrow focus was mentioned in contexts where word order marks the focused phrase only, but not a narrow focus within (such as in “Which small umbrella was lost yesterday?” “The small YELLOW umbrella.”, where the DP is in preverbal focus position in Hungarian, but narrow focus falls on the second adjective only). Intonational marking of narrow focus in such contexts was found in Langer & Kügler (2023). Similarly, in situations where postverbal focus is acceptable (as has been found in Buza, Wagner, & Jabeen, 2026), an amplified role of prosodic marking may result in coordination with f_0 . As the stable stressed syllable (or its nucleus) is available, however, in these lexical affiliates as well, it may be the case that co-speech gestures remain coordinated with that landmark.

As also discussed in relation to the role of information structure in coordination, distinct patterns may arise in more spontaneous gesture and language production where pragmatic considerations inform production organically. Although no study directly compares the alignment of speech and co-speech gesture landmarks across different information structure categories in spontaneously produced vs. elicited speech and gesture, tentative comparisons can

be made based on separate studies examining the same language with different methodologies. One set of examples are studies on (different varieties of) English, which vary greatly in the genre of speech they examine (lecture, public discussion, conversation, read speech), but which find with consistency that the relevant points of coordination are the pitch accented syllable – often the peak of the pitch accent or the f_0 maxima – and the target of the co-speech gesture (see e.g. Shattuck-Hufnagel & Ren, 2018 (lecture with spontaneous gesture); Krivokapic et al., 2016 (elicited speech and gesture); Loehr 2012 (conversation with spontaneous gesture); Mendoza-Denton & Jannedy, 2011 (public discussion with spontaneous gesture); Leonard & Cummins, 2011 (elicited speech and gesture); Yasinnik et al., 2004 (lecture with spontaneous gesture)). Given the relative consistency in the findings across studies, it is possible that the gestural landmark of coordination does not vary with the type or genre of speech the gesture accompanies, but whether this pattern holds for Hungarian remains to be established.

In addition to the question of landmarks, one further consideration pointed out in the discussion is the tightness of alignment between the landmarks (see Gregori & Kügler (2026) on this point), which may be modulated by relative pragmatic prominence as arising due to the demands of discourse in spontaneous and even semi-spontaneous speech. As *Chapter 3* indicates, such differences do not arise in Hungarian in read speech with elicited gestures, and whether they arise across information structural units in different genres of speech (with varying discourse demands on the expression of pragmatic prominence) remains to be seen.

A further question may arise in connection with the type of co-speech gesture examined. The present dissertation analyzed the coordination patterns of pointing / deictic gestures, which are a type of referential gesture. While the question of whether non-referential gestures behave similarly to referential ones in their coordination patterns remains open, note that the pattern of

coordination, i.e. which landmarks are involved, appears to hold regardless of the gestures' referential/non-referential nature in English, see Shattuck-Hufnagel & Ren (2018). Whether their kinematic properties (such as the temporal properties examined in *Chapter 4*) under prosodic boundaries are similar is to be examined in future studies as well (note however that Krivokapić et al. (2024) find both non-referential (beat) and referential (deictic) gestures to be equally loosely coordinated with prosodic structure in English; therefore, both types of gestures are predicted to lengthen under prosodic boundaries).

In a similar vein, it should be acknowledged that the dissertation only examined co-speech behavior in accompanying proper names. Future work may consider alignment and temporal patterns given more varied lexical affiliates, including other types of referential expressions, quantified expressions, and verb phrases. Given the observed role of prosodic structure, predictions need not necessarily change with the type of lexical affiliate: we may expect coordination with the nucleus of the stressed syllable (when present) in Hungarian under such circumstances as well. Note though that, taking quantified expressions as an example, the lexical affiliate may be the entire phrase (*minden párna*, 'every pillow') or some part of it (*minden párna*, 'pillow'), depending on discourse demands, and the co-speech gesture is predicted to coordinate with the corresponding landmark. With respect to prosodic boundaries, when present, we may expect them to have a lengthening effect regardless of the type of lexical affiliate (although the effect may be larger/smaller depending on the magnitude of the boundary, but this is as predicted by the π -gesture model).

As noted in *Chapter 3*, the data collected as part of the dissertation does not allow for the analysis of spontaneous language and gesture⁴ production, but it does allow for the future examination of spontaneous eyebrow, head, and torso movements across IS conditions, and although the question-answer pairs are elicited, the accompanying non-manual gestures constitute a spontaneously produced set. While their frequency of occurrence may be related to the presence of the elicited pointing gesture, their co-occurrence, alignment, and kinematic patterns could lead to further insight regarding the role of information structure in speech and co-speech gesture co-production, which may speak to the aforementioned questions as well.

Next, the effect of prosodic boundaries and prominence on the temporal properties of co-speech gestures can be further explored as well. Regarding the former, all else being equal, including a longer pre-target word could detangle the effects of the targeted boundary (*ma_Mima* vs. *Ma? Mima*) and preceding boundaries. As for the effect of prominence, temporal effects were not observed (no lengthening of the consonants and vowels as a function of information structural differences), and so the presence of a temporal modulation μ -gesture could not be inferred, but as suggested, spatial effects could be examined with the available data in the future to determine if a spatial modulation μ -gesture is present in marking prominence across modalities. As in the case of coordination, lengthening effects could be further explored in the set of spontaneously produced non-manual co-speech gestures as well.

Future work pursuing the directions outlined will further elucidate the interplay of prosody, information structure, and co-speech gestures.

⁴ Note that although the co-speech gesture was elicited in the study, participants were not restricted in the type and magnitude of the pointing gesture they produced, allowing for individual variability as well as a degree of spontaneity.

5.6 Conclusion

The dissertation has examined the respective roles prosody and information structure have in speech and deictic co-speech gesture alignment in Hungarian. In particular, the first study established the vowel target of the first, lexically stressed syllable and the target of the pointing gesture as the points of coordination in Hungarian. The stability of this coordination was not significantly affected by information structure, intonation contour, or their interaction. The second study examined if pointing gestures are affected by prosodic boundaries and prominence in the same way as speech gestures are, in order to determine whether or not co-speech gestures are influenced by prosodic structure. The hypotheses regarding lengthening were not borne out in the expected manner, but some evidence was found that boundaries lengthen gesture phases co-extensive with them. Prominence had no lengthening effect on speech gestures, but lengthening was observed in the case of pointing gestures, which was attributed to information structure modulating boundary magnitude, and thus pointing phase durations. Models of speech production can accommodate these findings via the linking of prosodic and gesture formulation modules (as suggested in Türk, 2020). Altogether, serving also as a mediator between information structure and gesturing, prosodic structure appears to drive the synchrony between speech and co-speech gesture production in Hungarian.

These findings largely support prior observations about gesture behavior in other languages, in that the co-speech gesture coordinates with the prominence-bearing syllable, although language specific patterns have also been noted, which likely relate to the prosodic typology of the languages so far studied. Future research will have to sharpen the emerging picture of gestural coordination.

Bibliography

- Abner, N., Cooperrider, K., & Goldin-Meadow, S. (2015). Gesture for linguists: A handy primer. *Language and linguistics compass*, 9(11), 437-451.
- Ahlmann-Eltze, C., & Patil, I. (2021). ggsignif: R Package for Displaying Significance Brackets for 'ggplot2'. PsyArxiv. doi:10.31234/osf.io/7awm6
- Ambrazaitis, G., & House, D. (2016). Multimodal levels of prominence: the use of eyebrows and head beats to convey information structure in Swedish news reading. In *ISGS Conference 2016. 7th Conference of the International Society for Gesture Studies*. Paris, France, July 18-22, 2016 (pp. 319-319). New Sorbonne University Paris 3.
- Ambrazaitis, G., & House, D. (2017). Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Communication*, 95, 100-113.
- Arvaniti, A., & Fletcher, J. (2020). The Autosegmental-Metrical Theory of Intonational Phonology. *The Oxford Handbook of Language Prosody*, 77-95.
- Azar, Z., Backus, A., & Özyürek, A. (2019). General-and language-specific factors influence reference tracking in speech and gesture in discourse. *Discourse Processes*, 56(7), 553-574.
- Baills, F., & Baumann, S. (2025). Prosody and head gestures as markers of information status in French as a native and foreign language. *Language and Cognition*, 17, e42. Bosker et al., 2026)

- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. convergence, 12(1), 2.
- Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- Baumann, S. (2016). "Contrast: Dissecting an elusive information-structural notion and its role in grammar." In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 239-260). Oxford University Press.
- Beckman, Mary E. and Janet B. Pierrehumbert. (1986). Intonational structure in English and Japanese. *Phonology Yearbook* 3: 255-310.
- Benus, Stefan; Reichel, Uwe D. und Mády, Katalin (2014): Modeling Accentual Phrase Intonation in Slovak and Hungarian. In: Veselovská, L. und Janebová, M. (eds.) : *Complex Visibles Out There. Olomouc modern language series*, Vol. 4. Olomouc, Czech Republic: Palacký University. pp. 677-689
- Bergmann, K., Aksu, V., & Kopp, S. (2011). The relation of speech and gestures: Temporal synchrony follows semantic synchrony. In *Proceedings of the 2nd Workshop on Gesture and Speech in Interaction* (GeSpIn 2011).
- Boersma, Paul & Weenink, David (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.06, retrieved from <https://praat.org>
- Bombien, L., Mooshammer, C., & Hoole, P. (2013). Articulatory coordination in word-initial clusters of German. *Journal of Phonetics*, 41(6), 546-561.
- Bosker, H. R., Hoetjes, M., Hustin, D., Pouw, W., & Van Maastricht, L. (2026). Foreign language learners show a kinematic accent in their co-speech hand movements. *Open Mind*, 10, 66-78.

- Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech* 49.451–493.
- Browman, CP., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6(2), 201-251.
- Browman, CP., & Goldstein LM. (1992). Articulatory phonology: an overview. *Phonetica* 49, 155–180. (doi:10.1159/000261913)
- Browman, CP., Goldstein L. (1995). Dynamics and articulatory phonology. In *Mind as motion: explorations in the dynamics of cognition* (eds RF Port, T Van Gelder), pp. 175–193. Cambridge, MA: The MIT Press.
- Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin de la communication parlée*, (5), 25-34.
- Buza, A., Wagner, P., & Jabeen, F. (2026). Anything goes? On the appropriateness of mismatches between prosody and syntax in Hungarian narrow focus marking. In *Proc. SpeechProsody 2026* (pp. 123-127).
- Bürling, D. (2003). On D-trees, beans, and B-accent. *Linguistics and philosophy*, 26, 511-545.
- Byrd, D., & Krivokapić, J. (2021). Cracking prosody in articulatory phonology. *Annual Review of Linguistics*, 7(1), 31-53.
- Byrd, D., & Saltzman, E. (1998). Intra-gestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, 26(2), 173-199.
- Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31(2), 149-180.

- Calhoun, S. (2010). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*, 25(7-9), 1099–1140.
- Canty, A., & Ripley, B. (2021). boot: bootstrap R (S-Plus) Functions.—R package ver. 1.3-28.
- Carignan, C., Esteve-Gibert, N., Løevenbruck, H., Dohen, M., & d'Imperio, M. (2024). Co-speech head nods are used to enhance prosodic prominence at different levels of narrow focus in French. *The Journal of the Acoustical Society of America*, 156(3), 1720-1733.
- Chafe, Wallace L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In: Charles N. Li (ed.): *Subject and topic*, 27–55. Academic Press, New York.
- Chernick, M. R. (1999). Bootstrap methods: A practitioner's guide (*Wiley series in probability and statistics*). Hoboken, NJ: Wiley.
- Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, 37(4), 466-485.
- Cho, T., Son, M., & Kim, S. (2016). Articulatory reflexes of the three-way contrast in labial stops and kinematic evidence for domain-initial strengthening in Korean. *Journal of the International Phonetic Association*, 46(2), 129-155.
- Cole, J., & Shattuck-Hufnagel, S. (2016). New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology*, 7(1).
- Cruz, M., Swerts, M., & Frota, S. (2015). Variation in tone and gesture within language. In *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Ćwiek, A., & Fuchs, S. (2025). The Scope and Limits of Iconic Prosody: Head Angle Predicts f0 Changes While Object Size Effects Are Absent. *Open Mind*, 9, 1959-1981.
- Davison A, Hinkley D (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge. ISBN 0-521-57391-2, doi:10.1017/CBO9780511802843.

- d'Imperio, M., & Cangemi, F. (2011). Phrasing, register level downstep and partial topic constructions in Neapolitan Italian. *Intonational phrasing in Romance and Germanic: Cross-linguistic and bilingual studies*, 10, 75-94.
- Debreslioska, S., & Gullberg, M. (2019). Discourse reference is bimodal: how information status in speech interacts with presence and viewpoint of gestures. *Discourse Processes*, 56(1), 41-60.
- Debreslioska, S., & Gullberg, M. (2020). What's new? Gestures accompany inferable rather than brand-new referents in discourse. *Frontiers in Psychology*, 11, 1935.
- Debreslioska, S., & Gullberg, M. (2022). Information status predicts the incidence of gesture in discourse: An experimental study. *Discourse Processes*, 59(10), 791-827.
- Debreslioska, S., Özyürek, A., Gullberg, M., & Perniss, P. (2013). Gestural viewpoint signals referent accessibility. *Discourse Processes*, 50(7), 431-456.
- De Ruiter, J. P. (2000). 14 The production of gesture and speech. *Lang. Gest.*, 2(284).
- De Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in cognitive science*, 4(2), 232-248.
- Destruel, Emilie, Daniel Velleman, Edgar Onea, Dylan Bumford, Jingyang Xue & David Beaver. (2015). A cross-linguistic study of the non-at-issueness of exhaustive inferences. In Florian Schwarz (ed.), *Experimental perspectives on presuppositions*, 135–156. Springer.
- Ebert, C., Evert, S., & Wilmes, K. (2011). Focus marking via gestures. In *Proceedings of Sinn und Bedeutung* (Vol. 15, pp. 193-208).

- Egbert, J., Plonsky, L. (2020). Bootstrapping Techniques. In: Paquot, M., Gries, S.T. (eds) *A Practical Handbook of Corpus Linguistics*. Springer, Cham. https://doi.org/10.1007/978-3-030-46216-1_24
- É. Kiss, Katalin. (1981). Structural relations in Hungarian, a "free" word order language. *Linguistic Inquiry*, 12(2), 185-213.
- É. Kiss, Katalin. (1994). Sentence structure and word order. In *The Syntactic Structure of Hungarian*, ed. by F. Kiefer and K. É. Kiss, 1–90. San Diego, London: Academic Press.
- É. Kiss, Katalin. (1998). Identificational focus versus information focus. *Language*, 74(2), 245-273.
- É. Kiss, Katalin. (2002). The syntax of Hungarian. *Cambridge syntax guides*. Cambridge, UK ; New York: Cambridge University Press.
- É. Kiss, Katalin. (2008). Topic and focus: Two structural positions associated with logical functions in the left periphery of the Hungarian sentence. *Acta linguistica Hungarica*, 55(3–4), 287–296. Akadémiai Kiadó.
- É. Kiss, Katalin. (2015). Negation in Hungarian. In *Negation in Uralic Languages* (pp. 219-238). John Benjamins.
- É. Kiss, Katalin. (2016). Discourse functions: The case of Hungarian. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 663-685). Oxford University Press.
- Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., & Prieto, P. (2017). The timing of head movements: The role of prosodic heads and edges. *The Journal of the Acoustical Society of America*, 141(6), 4727-4739.

- Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, 56(3), 850-865.
- Ferré, G. (2010). Timing relationships between speech and co-verbal gestures in spontaneous French. In *Language Resources and Evaluation, Workshop on Multimodal Corpora* (Vol. 6, pp. 86-91).
- Ferré, G. (2014). A multimodal approach to markedness in spoken French. *Speech Communication*, 57, 268-282.
- Fox, J., & Weisberg, S. (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Fletcher, J. (2010). The prosody of speech: Timing and rhythm. *The handbook of phonetic sciences*, 521-602.
- Franich, K. (2026). Variation in the Fine Timing of Co-Speech Gestures Among Bilingual Speakers. In *Proc. Speech Prosody 2026* (pp. 88-92).
- Franich, K., Keupdjio, H., & Nwosu, V. (2025). The timing of speech and gesture in two Niger-Congo languages: Implications for word-level prominence. *Glossa: a journal of general linguistics*, 10(1).
- Fuchs, S., & Reichel, U. D. (2016). On the relationship between pointing gestures and speech production in German counting out rhymes: Evidence from motion capture data and speech acoustics. *Proceedings of P&P*, 12, 1-4.
- Fung, H. S. H., & Mok, P. P. K. (2018). Temporal coordination between focus prosody and pointing gestures in Cantonese. *Journal of Phonetics*, 71, 113-125.

- Galati, A., & Brennan, S. E. (2014). Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture. *Language, Cognition and Neuroscience*, 29(4), 435-451.
- Gao, M. (2008). *Mandarin tones: An articulatory phonology account* [Ph. D. thesis]. Yale Univ New Haven, CT.
- Genzel, S., Ishihara, S., & Surányi, B. (2015). The prosodic expression of focus, contrast and givenness: A production study of Hungarian. *Lingua*, 165, 183-204.
- Gerőcs, Mátyás, Anna Babarczy and Balázs Surányi. (2014). Exhaustivity in focus: Experimental evidence from Hungarian. In J. Emonds and M. Janebová (eds.) *Language use and linguistic structure*. Olomouc: Palacký University. 181–194.
- Gerwing, Jennifer (2003). *The effect of immediate communicative function on the physical form of conversational hand gestures*. Unpublished Masters thesis, University of Victoria, Victoria, British Columbia.
- Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture*, 4(2), 157-195.
- Goldstein L, Byrd D, Saltzman E. (2006) The role of vocal tract gestural action units in understanding the evolution of phonology. In *Action to language via the mirror neuron system* (ed. M Arbib), pp. 215–249. New York, NY: Cambridge University Press.
- Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. *Frontiers in phonetics and speech science*, pp-239.
- Gósy, Mária and Krepsz, Valéria (2018) Phrase-final Lengthening of Phonemically Short and Long Vowels in Hungarian Spontaneous Speech Across Ages. In: *CHALLENGES IN*

- ANALYSIS AND PROCESSING OF SPONTANEOUS SPEECH*. MTA Nyelvtudományi Intézet, Budapest, pp. 99-126. ISBN 978-963-9074-76-7 (online)
- Götze, M., Weskott, T., Endriss, C., Fiedler, I., Hinterwimmer, S., Petrova, S., ... & Stoel, R. (2007). Information structure. *Interdisciplinary studies on information structure: ISIS*, (7), 147-187.
- Gregori, A., & Kügler, F. (2026). Temporal alignment of prosody and gesture in focus: Unimodal and Multimodal Hyperarticulation. *Journal of Phonetics*, 117, 101503.
- Gregori, A., Sánchez-Ramón, P. G., Prieto, P., & Kügler, F. (2024). Prosodic and gestural marking of focus types in Catalan and German. In *Proc. SpeechProsody 2024* (pp. 891-895).
- Grice M (2006), Intonation. In: Keith Brown, (Editor-in-Chief) *Encyclopedia of Language & Linguistics*, Second Edition, volume 5, pp. 778-788. Oxford: Elsevier.
- Gyuris, B. (2002). *The semantics of contrastive topics in Hungarian* (Doctoral dissertation). Eötvös Loránd University, Budapest
- Gyuris, B. (2009). The semantics and pragmatics of the contrastive topic in Hungarian. *Lexica*.
- Gyuris, B. (2012). The information structure of Hungarian. *The expression of information structure*, 159-186.
- Gyuris, B., & Mády, K. (2014). Contrastive topics between syntax and pragmatics in Hungarian: an experimental analysis. In *Proceedings from the annual meeting of the Chicago Linguistic Society* (Vol. 46, No. 1, pp. 147-162). Chicago Linguistic Society.
- Hedberg, N. 2003. The prosody of contrastive topic and focus in spoken English. In *Pre-proceedings of the workshop on information structure in context*, 14–52. Stuttgart: IMS

- Heim, I. (2002). File change semantics and the familiarity theory of definiteness. *Formal semantics: The essential readings*, 223-248.
- Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., & Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language*, 79, 1-17.
- Höhle, Tilman (1988). "Vorwort und Nachwort zu 'Verum-Fokus'." In: *Sprache und Pragmatik. Arbeitsberichte 5*. Lund: Germanistisches Institut der Universität Lund, 1-7.
- Höhle, Tilman (1992). "Über Verum-Fokus im Deutschen." In: *Informationsstruktur und Grammatik. Linguistische Berichte Sonderhefte*, 4, 112-141.
- Holler, J., Bavelas, J., Woods, J., Geiger, M., & Simons, L. (2022). Given-new effects on the duration of gestures and of words in face-to-face dialogue. *Discourse Processes*, 59(8), 619-645.
- Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of language and social psychology*, 26(1), 4-27.
- Holler, J., Tutton, M., & Wilkin, K. (2011). Co-speech gestures in the process of meaning coordination. In *2nd GESPIN-Gesture & Speech in Interaction*. Bielefeld, 5-7 Sep 2011.
- Im, S., & Baumann, S. (2020). Probabilistic relation between co-speech gestures, pitch accents and information status. *Proceedings of the Linguistic Society of America*, 5(1), 685-697.
- Isosävi, J., & Vecsernyés, I. (2022). Addressing, greeting and related gestures in the opening sequences of Finnish, French and Hungarian YouTube videos. *Contrastive Pragmatics*, 3(3), 363-396.
- Jacobs, J. (2001). The dimensions of topic-comment. *Linguistics*, 39(4).

- Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language*, 56(2), 291-303.
- Jun, S.-A. (2005). Prosodic Typology. In S.-A. Jun (Ed.), *Prosodic typology: The Phonology of intonation and phrasing* (pp. 410-458). Oxford University Press.
- Jun, S.-A. (2014). Prosodic Typology: By prominence type, word prosody, and macro-rhythm. In S.-A. Jun (Ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing* (pp. 520-539). Oxford University Press.
- Jun, S. -A. (2025). Prosodic typology: Intonational tone types and functions. *Contemporary linguistics: Integrating languages, communities, and technologies*, 7, 93-111.
- Kadavá et al. (2023). Is Gesture-Speech Physics At Work In Rhythmic Pointing? Evidence From Polish Counting-Out Rhymes. In: Radek Skarnitzl & Jan Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 4190-4194). Guarant International
- Kaland, Constantijn. (2026). Contour Clustering (Version 2026-02) [Computer software].
<https://constantijnkaland.github.io/contourclustering/>
- Kaland, Constantijn. (2023). Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association* 53(1), 159-188. doi:10.1017/S0025100321000049 [pdf]
- Kassambara A (2023). `_ggpubr: 'ggplot2' Based Publication Ready Plots_`. R package version 0.6.0, <<https://CRAN.R-project.org/package=ggpubr>>.
- Katsika, A. (2012). *Coordination of prosodic gestures at boundaries in Greek*. Yale University.
- Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics*, 44, 62-82.

- Káldi, Tamás, & Babarczy, Anna. (2018). Linguistic exhaustivity inference is context dependent: A visual-world eye-tracking study on Hungarian focus. *Acta linguistica academica*, 65(4), 547–595. Akadémiai Kiadó.
- Káldi, Tamás, Babarczy, Anna & Bende-Farkas, Ágnes (2017). Hungarian Focus: Presuppositional Content and Exhaustivity Revisited. In J. Emonds and M. Janebová (Eds.), *Language Use and Linguistic Structure: Proceedings of the Olomouc Linguistics Colloquium 2016*. (pp. 245-262). Palacký University, Olomouc.
- Kálmán, L. & Nádasdy, Á. (1994). A hangsúly . [Stress .] In *Strukturális magyar nyelvtan 2. Fonológia. [A structural grammar of Hungarian. 2. Phonology.]*, Ferenc Kiefer (ed .), 393–467 . Budapest: Akadémiai Kiadó
- Kendon, A. (1972). Some relationships between body motion and speech. An analysis of an example. In A. Siegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 170–210). Elmsford, NY: Pergamon Press.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. *The relationship of verbal and nonverbal communication*, 25(1980), 207-227.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge University Press.
- Kim, J., Cvejic, E., & Davis, C. (2014). Tracking eyebrows and head gestures associated with spoken prosody. *Speech Communication*, 57, 317-330.
- Kiss, J. (2006). Nyelvjárások, regionális nyelvváltozatok. [Dialects, regional language varieties.] In F. Kiefer. (Ed.). *Magyar nyelv. [Hungarian language.]* (517-548). Budapest: Akadémiai Kiadó.
- Kita, Sotaro. 1990. *The Temporal Relationship between Gesture and Speech: A Study of Japanese-English Bilinguals*. MA thesis, University of Chicago.

- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and language*, 48(1), 16-32.
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica* (Since 2017 *Acta Linguistica Academica*), 55(3-4), 243-276.
- Krifka, M., & Musan, R. (2012). Information structure: Overview and linguistic issues. In M. Krifka & R. Musan (Eds.), *The Expression of Information Structure* (pp. 1-44). De Gruyter Mouton. <https://doi.org/10.1515/9783110261608.1>
- Krivokapić, Jelena. (2014). "Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes." *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1658: 20130397.
- Krivokapić, J., Styler, W., & Byrd, D. (2022). The role of speech planning in the articulation of pauses. *The Journal of the Acoustical Society of America*, 151(1), 402-413.
- Krivokapic, J., Tiede, M., Tyrone, M., & Goldenberg, D. (2016). Speech and manual gesture coordination in a pointing task. In *Proc. Speech Prosody 2016* (pp. 1240-1244).
- Krivokapić, J., Tiede, M. K., & Tyrone, M. E. (2017). A kinematic study of prosodic structure in articulatory and manual gestures: Results from a novel method of data collection. *Laboratory phonology*, 8(1).
- Krivokapić, J., Tiede, M., Tyrone, M. E., Purse, R., & Seo, J. (2024). The role of prosodic structure in the planning of coordinated speech and manual gestures. In *Proceedings of the 12th International Conference on Speech Prosody* (pp. 2-5).
- Kučerová, Ivona. 2007. *The syntax of givenness*. Massachusetts Institute of Technology

- Kügler, F., & Calhoun, S. (2020). "Prosodic encoding of information structure: A typological perspective." In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Language Prosody* (pp. 454-467). Oxford University Press.
- Kügler, F & Gregori, A. (2023). "Iconic Gestures in Focus – Synchronization of Prosody and Gestures in Prominence." In: Skarnitzl, R. & Volín, J. (eds) *Proceedings of 20th ICPhS 2023*, Prague, Czech Republic August 2023. 4125-4129 (ID 232). Guarant International.
- Kügler, F., Langer, C., Arnhold, A., & Ots, N. (2026). Pre-focal intonational variation in Hungarian as prominence reduction. In *Proc. SpeechProsody 2026* (pp. 58-62).
- Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.
- Ladd, D. R., & Arvaniti, A. (2023). Prosodic prominence across languages. *Annual Review of Linguistics*, 9(1), 171-193.
- Langer, C., & Kügler, F. (2021). Focus and Prosodic Cues in Hungarian Noun Phrases. In *Proc. 1st International Conference on Tone and Intonation (TAI)* (pp. 219-223).
- Langer, C., & Kügler, F. (2023) "Prosody-Syntax Interface and Hungarian Noun Phrases". In: *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS20)*. Ed. by R.: Skarnitzl and J. Volín. Prague: Guarant International, pp. 1290–1294.
- Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26(10), 1457-1471.
- Lehečková, E., Jehlička, J., & Králová Zíková, M. (2022). Multimodal marking of information structure: gesture-prosody alignment across languages. *Linguistica Pragensia*, 32(1), 19-38.
- Lenth R (2023). *_emmeans: Estimated Marginal Means, aka Least-Squares Means_*. R package version 1.8.4-1, <<https://CRAN.R-project.org/package=emmeans>>.

- Levelt, W. J. M. (1989). *Speaking: from Intention to Articulation*. Cambridge: MIT Press.
- Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Netherlands: Springer. https://doi.org/10.1007/978-94-009-2037-8_16.
- Loehr, D. P. (2004). *Gesture and intonation*. Georgetown University.
- Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, 7(2), 179-214.
- Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory phonology*, 3(1), 71-89.
- Long, J. A. (2024). `_interactions`: Comprehensive, User-Friendly Toolkit for Probing Interactions_. doi:10.32614/CRAN.package.interactions
<<https://doi.org/10.32614/CRAN.package.interactions>>, R package version 1.2.0,
<<https://cran.r-project.org/package=interactions>>.
- Luchkina, T., & Cole, J. S. (2017). Structural and referent-based effects on prosodic expression in Russian. *Phonetica*, 73(3-4), 279-313.
- Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172.
- Markó, A., Bartók, M., Csapó, T. G., Deme, A., & Grácsi, T. E. (2019). The effect of focal accent on vowels in Hungarian: articulatory and acoustic data. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (Eds.) *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2715-2719). Australasian Speech Science and Technology Association Inc., and International Phonetic Association.

- Markó, A., Bartók, M., Grácz, T.E., Deme, A., Csapó, T.G. (2018) Prominence Effects on Hungarian Vowels: A Pilot Study. *Proc. Speech Prosody 2018*, 868-872, doi: 10.21437/SpeechProsody.2018-175
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.
- Mády, K. (2012). Deaccentuation in Hungarian and its logical background. In: *6th Speech Prosody Conference*, Shanghai, China.
- Mády, K. (2015). Prosodic (non-) realisation of broad, narrow and contrastive focus in Hungarian: a production and a perception study. In *Proc. Interspeech 2015*, 948-952, doi: 10.21437/Interspeech.2015-18
- Mády, K., & Kleber, F. (2010). Variation of pitch accent patterns in Hungarian. In *Proc. Speech Prosody 2010*, paper 924, doi: 10.21437/SpeechProsody.2010-199
- Mády, K., Kleber, F., Reichel, U. D., & Szalontai, A. (2016). The interplay of prominence and boundary strength: a comparative study. In *Proc. Phonetik und Phonologie*, Munich, Germany.
- Mády, K., Reichel, U. D., & Beňuš, Š. (2013). Accentual phrase in languages with fixed word stress: a study on Hungarian and Slovak. In *Workshop Advancing Prosodic Transcription for Spoken Language Science and Technology II, Phonetics and Phonology in Iberia 2013*.
- Mády, K., Reichel, U. D., Kohári, A., Molnár, C. S., & Szalontai, Á. (2023). Prosodic cues of distinguishing neutral and non-neutral yes/no questions in Hungarian: The acoustics of surprise. In *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 1608-1612).

- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago Press.
- McNeill, D. (2000). *Growth points in thinking-for-speaking*. Language and Gestures/Cambridge University Press.
- McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.
<http://doi.org/10.7208/chicago/9780226514642.001.0001>
- McNeill, D. (2006). Gesture: a psycholinguistic approach. In *The Encyclopedia of Language and Linguistics* (Amsterdam & Boston: Elsevier), 58–66
- Mehlhorn, G. 2001. Produktion und Perzeption von Hutkonturen im Deutschen. *Linguistische Arbeitsberichte* 77.31–57.
- Mendoza-Denton, N., & Jannedy, S. (2011). Semiotic layering through gesture and intonation: A case study of complementary and supplementary multimodality in political speech. *Journal of English Linguistics*, 39(3), 265-299.
- Mol, L., & Kita, S. (2012). Gesture structure affects syntactic structure in speech. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34, No. 34).
- Molnár, Valéria (2006). ‘On Different Kinds of Contrast’, in V. Molnár and S. Winkler (eds), *The Architecture of Focus*. Berlin: Mouton de Gruyter, 197–234.
- Mücke, D., Nam, H., Hermes, A., & Goldstein, L. (2012). Coupling of tone and constriction gestures in pitch accents. *Consonant clusters and structural complexity*, 26, 205.
- Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation—Is it mediated by accentuation?. *Journal of Phonetics*, 44, 47-61.
- Mycock, L. (2010). Prominence in Hungarian: the prosody–syntax connection. *Transactions of the Philological Society*, 108(3), 265-297.

- Myslyvyi, K., Lorenzen, J., & Baumann, S. (2026). Multimodal prosody in German—Evidence for the additive nature of gestural and accentual prominence in form and timing. In *Proc. SpeechProsody 2026* (pp. 559-563).
- Nagy, L. C., & Császi, I. N. (2015). *Magyar nyelvjárások. [Hungarian dialects.]* Tinta Könyvkiadó.
- Nam, H., Goldstein, L. and Saltzman, E. (2009). Self-organization of syllable structure: a coupled oscillator model. In F. Pellegrino, E. Marisco and I. Chitoran, (eds.). *Approaches to phonological complexity*. Berlin, New York: Mouton de Gruyter, 299–328.
- Nam, H., & Saltzman, E. (2003, August). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th international congress of phonetic sciences* (Vol. 1, No. 2253-2256).
- Nolan, F. (2022). The Rise and Fall of the British School of Intonation Analysis. In Barnes, J. & Shattuck-Hufnagel, S. (Eds.), *Prosodic Theory and Practice* (pp. 319-349). The MIT Press
- Odden, D. (1986). *Phonological Theory: The Case of English*. Academic Press.
- Owoyele, B., Trujillo, J., De Melo, G., & Pouw, W. (2022). Masked-Piper: Masking personal identities in visual recordings while preserving multimodal information. *SoftwareX*, 20, 101236.
- Paoli, Sandra (2009). ‘Contrastiveness and New Information. A New View on Focus’, *Rivista di Grammatica Generativa* 34: 137–161.
- Parrell, B., Goldstein, L., Lee, S., & Byrd, D. (2014). Spatiotemporal coupling between speech and manual motor actions. *Journal of phonetics*, 42, 1-11.

- Partee, B. H. (1999). Focus, quantification, and semantics-pragmatics issues. *Focus: Linguistic, cognitive, and computational perspectives*, 213.
- Paschen, L., Fuchs, S., & Seifart, F. (2022). Final lengthening and vowel length in 25 languages. *Journal of Phonetics*, 94, 101179.
- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation. MIT).
- Pierrehumbert, J., & Beckman, M. (1986). Intonational structure in Japanese and English. *Phonology yearbook*, 3, 255-310.
- Pinheiro J, Bates D, R Core Team (2022). *_nlme: Linear and Nonlinear Mixed Effects Models_*. R package version 3.1-157, <<https://CRAN.R-project.org/package=nlme>>.
- Pouw, W., & Fuchs, S. (2022). Origins of vocal-entangled gesture. *Neuroscience & Biobehavioral Reviews*, 141, 104836.
- Pouw, W., Harrison, S. J., & Dixon, J. A. (2020a). Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony. *Journal of Experimental Psychology: General*, 149(2), 391.
- Pouw, W., Harrison, S. J., & Dixon, J. A. (2020b). Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures. PsyArXiv Preprints. <https://doi.org/10.31234/osf.io/xv78f>
- Pouw, W., Harrison, S. J., & Dixon, J. A. (2020c). Gesture-speech physics in fluent speech and rhythmic upper limb movements. PsyArXiv Preprints. <https://doi.org/10.31234/osf.io/y9k2m>

- Pouw, W., Paxton, A., Harrison, S. J., & Dixon, J. A. (2020). Acoustic information about upper limb movement in voicing. *Proceedings of the National Academy of Sciences*, 117(21), 11364-11367.
- Pouw, W., Werner, R., Burchardt, L. S., & Selen, L. P. (2025). The human voice aligns with whole-body kinetics. *Proceedings of the Royal Society B: Biological Sciences*, 292(2047).
- Prieto, P., Esteve-Gibert, N., & Shattuck-Hufnagel, S. (2025). Towards a novel conceptualization of prosody that accounts for spoken and visual signals: The modality-neutral prosodic framework hypothesis. *Gesture*, 23(1-2), 119-159.
- R Core Team (2025). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Reichel, U. D., Mády, K., & Kleber, F. (2016). How prominence and prosodic phrasing interact. In: Jokisch, O. (Ed.): *Elektronische Sprachverarbeitung. Studentexte zur Sprachkommunikation* 81, pp 153-159, TUDpress, Dresden
- Reinhart, T. (1981). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, 27.
- Repp, S. (2016). "Contrast: Dissecting an elusive information-structural notion and its role in grammar." In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 239-260). Oxford University Press.
- Roberts, C. (1996.) Information structure in discourse: towards an integrated formal theory of pragmatics. In Jae-Hak Yoon & Andreas Kathol (eds.), *Papers in semantics (Working Papers in Linguistics 49)*. The Ohio State University.

- Rochet-Capellan, A., Laboissière, R., Galván, A., & Schwartz, J. L. (2008). The speech focus position effect on jaw–finger coordination in a pointing task. *Journal of Speech, Language, and Hearing Research*, 51(6), 1507-1521.
- Rooth, Mats 1992. A theory of focus interpretation. In: *Natural Language Semantics* 1:75–116.
- Roustan, B. & Dohen, M. (2010). Co-production of contrastive prosodic focus and manual gestures: Temporal coordination and effects on the acoustic and articulatory correlates of focus. *Speech Prosody 2010*, 100110, 1–4.
- Rohrer, P. (2022). *A temporal and pragmatic analysis of gesture-speech association: A corpus-based approach using the novel MultiModal MultiDimensional (M3D) labeling system* (Doctoral dissertation, Nantes Université; Universitat Pompeu Fabra (Barcelone, Espagne)).
- Rohrer, P. L., Delais-Roussarie, E., & Prieto, P. (2023). Visualizing prosodic structure: Manual gestures as highlighters of prosodic heads and edges in English academic discourses. *Lingua*, 293, 103583.
- Rohrer, P. L., Hong, Y., & Bosker, H. R. (2024). Gestures time to vowel onset and change the acoustics of the word in Mandarin. In *Speech Prosody 2024* (pp. 866-870).
- Rohrer, P. L., Prieto, P., & Delais-Roussarie, E. (2019, August). Beat gestures and prosodic domain marking in French. In *19th International Congress of Phonetic Sciences* (pp. 1500-1504).
- Roustan, B. & Dohen, M. (2010). Co-production of contrastive prosodic focus and manual gestures: Temporal coordination and effects on the acoustic and articulatory correlates of focus. *Speech Prosody 2010*, 100110, 1–4.

- Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2014). Effects of perturbation and prosody on the coordination of speech and gesture. *Speech Communication*, 57, 283-300.
- Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008, May). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th international conference on speech prosody (Speech Prosody 2008)*, Campinas, Brazil (pp. 175-184).
- Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, 19(4), 499-526.
- Selkirk, Elisabeth (1995). Sentence prosody: intonation, stress and phrasing. In J. Goldsmith (ed.) *The handbook of phonological theory*. Cambridge, MA: Blackwell. 550-569.
- Shattuck-Hufnagel, S., & Prieto, P. (2019, August). Dimensionalizing co-speech gestures. In *Proceedings of the international Congress of phonetic sciences* (Vol. 5, pp. 1490-1494).
- Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology*, 9, 1514.
- Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25, 193-247.
- Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62:626-633.
- Siptár, P., & Törkenczy, M. (2000). *The phonology of Hungarian*. OUP Oxford.

- Schegloff, Emanuel A. (1984). On some gestures' relation to talk. In J. M. Atkinson & J. Heritage (eds.), *Structures of Social Action*, pp. 266–298. Cambridge: Cambridge University Press.
- Schiffman, H. F. (2017). Diglossia as a sociolinguistic situation. *The handbook of sociolinguistics*, 205-216.
- Schlenker, P. (2018). Gesture projection and cosuppositions. *Linguistics and Philosophy*, 41(3), 295-365.
- Schlenker, P. (2020). Gestural grammar. *Natural Language & Linguistic Theory*, 38(3), 887-936.
- Schwarzschild, R. (1999). GIVENness, AvoidF and other constraints on the placement of accent. *Natural language semantics*, 7(2), 141-177.
- Stalnaker, Robert (1978). "Assertion". In Cole, P (ed.). *Syntax and Semantics*, Vol. IX: Pragmatics. Academic Press.
- Stavropoulou, P., & Baltazani, M. (2021). The prosody of correction and contrast. *Journal of Pragmatics*, 171, 76-100.
- Surányi, B. (2016). Discourse-configurationality. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (Vol. 1). Oxford University Press.
- Swerts, M., & Krahmer, E. (2010). Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions. *Journal of Phonetics*, 38(2), 197-206.
- Szabolcsi, A. (1981). The semantics of topic-focus articulation. In *Formal Methods in the Study of Language, Part 1.*, ed. by J. Groenendijk, Th. Janssen, and M. Stokhof, 1–28. Amsterdam: Mathematisch Centrum.

- Szalontai, Á., & Surányi, B. (2020). Word order effects of givenness in Hungarian: Syntax or prosody?. In *Approaches to Hungarian* (pp. 137-164). John Benjamins.
- Szalontai, Á., Wagner, P., Mády, K., & Windmann, A. (2016). Teasing apart lexical stress and sentence accent in Hungarian and German. *Proc. Phonetik Und Phonologie Im Deutschsprachigen Raum* (2016), pp. 215-218
- Szendrói, Kriszta. (2003). A Stress-Based Approach to the Syntax of Hungarian Focus. *Linguistic review*, 20(1), 37–78. Berlin: Mouton de Gruyter.
- Tatár, C., Keshet, E., Krivokapić, J. (2026) Examining the role of prosody and information structure in Hungarian speech and co-speech gestural coordination: An EMA study. In *Proc. Speech Prosody 2026*, 496-500, doi: 10.21437/SpeechProsody.2026-100
- Tilsen, S. (2009). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive science*, 33(5), 839-879.
- Turk, A., Zhang, B., Li, K., & Flecha-Garcia, M. (2026). Do walking and talking entrain in a simultaneous walking+ talking task?. In *Proc. Speech Prosody 2026* (pp. 83-87).
- Türk, O. (2020). *Gesture, prosody and information structure synchronisation in Turkish* (Doctoral dissertation, Open Access Te Herenga Waka-Victoria University of Wellington).
- Türk, O. & Calhoun, S., (2023). Multimodal cues to intonational categories: Gesture apex coordination with tonal events. *Laboratory Phonology*, 14(1).
- Türk, O., & Calhoun, S. (2024). Phrasal synchronization of gesture with prosody and information structure. *Language and Speech*, 67(3), 702-743.

- van Heuven, V. J., & Turk, A. (2020). Phonetic correlates of word and sentence stress. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody* (pp. 150-165). Oxford University Press.
- Varga, László (1981). A topicról és a fókusz utáni elemek sorrendjéről [On the topic and the order of elements after the focus]. *Magyar Nyelv* 27 (pp. 198-200).
- Varga, L. (1983). Hungarian sentence prosody: an outline. *Folia Linguistica*, v.17, n. 1/2, p. 117
- Varga, L. (2002). *Intonation and stress: evidence from Hungarian*. Springer.
- Varga, László. 2010. Boundary tones and the lack of intermediate phrase in Hungarian (Revisiting the Hungarian Calling Contour). *The Even Yearbook* 9.
- Varga, L. (2017). The intonation of topic and comment in the Hungarian declarative sentence. *Finno-Ugric Languages and Linguistics*, 5(2).
- Vogel, I., Athanasopoulou, A., Pincus, N. (2015). Acoustic properties of prominence in Hungarian and the Functional Load Hypothesis. *Approaches to Hungarian*, 14, 267-292.
- Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech communication*, 57, 209-232.
- Werner, R., Selen, L., & Pouw, W. (2024). Arm movements increase acoustic markers of expiratory flow. *bioRxiv*, 2024-01.
- White, L., & Mády, K. (2008, May). The long and the short and the final: Phonological vowel length and prosodic timing in Hungarian. In *4th Speech Prosody Conference*, Campinas, Brasil (pp. 363-366).
- Wickham H, Hester J, Bryan J (2023). *_readr: Read Rectangular Text Data_*. R package version 2.1.4, <<https://CRAN.R-project.org/package=readr>>.

- Wickham H, Bryan J (2023). `_readxl: Read Excel Files_`. R package version 1.4.2, <<https://CRAN.R-project.org/package=readxl>>.
- Wickham H, François R, Henry L, Müller K, Vaughan D (2023). `_dplyr: A Grammar of Data Manipulation_`. R package version 1.1.2, <<https://CRAN.R-project.org/package=dplyr>>.
- Wickham H, Vaughan D, Girlich M (2023). `_tidyr: Tidy Messy Data_`. R package version 1.3.0, <<https://CRAN.R-project.org/package=tidyr>>.
- Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- Wickham H (2022). `_stringr: Simple, Consistent Wrappers for Common String Operations_`. R package version 1.5.0, <<https://CRAN.R-project.org/package=stringr>>.
- Yasinnik, Y., Renwick, M. & Shattuck-Hufnagel, S. 2004. The timing of speech- accompanying gestures with respect to prosody. *Proceedings of the International Conference: From Sound to Sense: +50 Years of Discoveries in Speech Communication*, MIT, Cambridge, 10-13 June, C97 – C102.
- Zou, Z., Arnhold, A. (2026) Prosody-syntax trade-offs in Mandarin clefts. *Proc. Speech Prosody 2026*, 108-112, doi: 10.21437/SpeechProsody.2026-22

ProQuest Number: 32992231

INFORMATION TO ALL USERS

The quality and completeness of this reproduction is dependent on the quality and completeness of the copy made available to ProQuest.



Distributed by
ProQuest LLC a part of Clarivate (2026).
Copyright of the Dissertation is held by the Author unless otherwise noted.

This work is protected against unauthorized copying under Title 17,
United States Code and other applicable copyright laws.

This work may be used in accordance with the terms of the Creative Commons license or other rights statement, as indicated in the copyright statement or in the metadata associated with this work. Unless otherwise specified in the copyright statement or the metadata, all rights are reserved by the copyright holder.

ProQuest LLC
789 East Eisenhower Parkway
Ann Arbor, MI 48108 USA