

Discontinuity as prosody: meaning and form of jump cuts on YouTube

Author: Maria Esipova, University of Konstanz

Abstract In this paper, I discuss the meaning and form of jump cuts, i.e., instances of visual discontinuity within a single video shot, in YouTube videos. I report the findings of a small-scale qualitative study, in which I sought (i) to identify the various semantic/pragmatic functions of the jump cut in YouTube videos, drawing comparisons with segmental, suprasegmental, and/or gestural realizations of similar meanings; and (ii) to analyze prosodic integration of sub-ip jump cuts with the speech stream, comparing it to other types of similar cross-channel integration. I concluded that integration of jump cuts with the speech stream draws on existing patterns of integration of segmental, suprasegmental, and gestural material into a coherent multi-channel signal at the level of both meaning and form—which allows jump cuts (and editing more broadly) to become a synergistic part of this multi-channel signal in YouTube videos.

1 Introduction

A jump cut (JC) in video editing is when a single shot is split into two parts, with some sort of visual discontinuity introduced between them. For instance, objects can change position within the frame, and/or the framing of the shot itself can change. This creates an abrupt, even jarring effect, which can be used intentionally to convey a range of meanings. One medium where JCs are used especially extensively is YouTube, which is a particularly interesting case from a linguistic perspective, as editing in YouTube videos integrates tightly with speech, both at the level of meaning and at the level of form, especially prosody. Looking at this integration can thus be a valuable contribution to the “Super Linguistics” research program, which seeks to apply the toolkit and the mindset of a linguist beyond language proper, with one of the ultimate goals being to achieve a better understanding of the universals of human cognition (see, e.g., Patel-Grosz et al. 2023).

Olson (2017) claims that the JC “has become ingrained in the basic visual language of [YouTube]” and attributes this to both practical and semantic considerations. From the practical perspective, intentional JCs can be used to conceal disfluencies or transitions between different takes: while some level of discontinuity is inevitable in these cases, a clearly intentional JC, with an obvious discontinuity, will appear less jarring than a seemingly unintentional one, with a less

ostensible discontinuity, which would just look like an editing error. Besides, most YouTubers are limited in their inventory of editing techniques for technical reasons (for instance, they often use a static single camera set-up), so JCs are one of the few tools available to them to make their videos appear more dynamic.

Aside from these practical considerations, Olson credits the popularity of the JC on YouTube to its “high degree of semiotic flexibility”. However, he himself only discusses a few of the JC’s functions, namely, “break[ing] up long, complex ideas into smaller, more manageable bites” and marking parenthetical statements and/or jokes, as in (1). Besides, in all examples he uses, including (1), JCs occur at clause boundaries in the speech stream and, thus, at intonational phrase (IP) or at least intermediate phrase (ip) boundaries in ToBI (Beckman & Ayers 1997) terms.

(1) This can serve to properly differentiate a joke from a serious statement by moving either very

far from or very close to the camera or,  <JC-side> [IP going back

to my own example]  <JC-back>, it can create an aside by literally moving aside. (Olson 2017; ‘aside’)

In this paper, I report the findings of a small-scale qualitative study of the JC in YouTube videos, seeking to provide a more detailed (albeit still not fully comprehensive) description of its meaning and form. In this study, I sampled and annotated 160 JC tokens (section 2). I used this data set (i) to identify the major semantic/pragmatic functions of the JC, drawing comparisons with segmental, suprasegmental, and/or gestural realizations of similar meanings (section 3); and (ii) to analyze prosodic integration of sub-ip JCs with the speech stream, comparing it to other types of similar cross-channel integration (section 4). I concluded that there are remarkable similarities between JCs and (near-)linguistic material at the level of both meaning and form, and that editing in YouTube videos can thus synergistically add to the speech stream.

2 Methodology

I selected 12 YouTube videos from 2018–2023 in 3 different genres (video essay; commentary; edutainment), with 2 channels per genre and 2 videos per channel. I sampled 10–20 tokens of what I judged to be intentional JCs or sequences of connected JCs from each video. A JC sequence can be a pair of JCs separating out a piece of the audiovisual signal (for instance, a parenthetical; see

subsection 3.2), a sequence of JCs separating list items (see subsection 3.3), or an instance of what I call a “ramp-up sequence” (see subsection 3.9). The total number of tokens was 160. I transcribed and annotated these tokens, categorizing them with respect to the function and form of the JC/JC sequence, using an inventory of tags that I developed and adjusted in the process of analyzing the data set. The list of source videos is provided at the end of this paper, and video clips of examples from this paper, alongside the spreadsheet with the transcribed and annotated tokens can be found at <https://tinyurl.com/jc-ex-nyj>.

I treated any form of abrupt visual discontinuity within a shot as a JC, and I used my own intuitions to judge if it was introduced intentionally, based primarily on how ostensible the discontinuity is. Of course, it is entirely possible that in some cases, a given JC was introduced primarily for practical reasons, with the discontinuity exaggerated to appear intentional. That said, at least in my data set, there didn’t seem to be many—if any—cases of a clearly intentional JC that wouldn’t be able to serve some identifiable function, so it would seem that even when creators use JCs for practical reasons, they typically do so at places where a JC would seem natural and meaningful.

I would also like to note that the vast majority of JC instances in my data set are abrupt frame shifts, not teleporting JCs like in (1). This makes sense from a practical standpoint, because JCs like in (1), or JCs where the background location changes (which are also discussed in Olson 2017), need to be planned during the filming process. Based on my data set and personal long-term experience of consuming YouTube content, as well as knowledge about how most creators film their videos (at least within the three genres I looked at), the vast majority of JCs in YouTube videos are actually introduced during the editing process. One interesting consequence of this is that the vast majority of JC instances in my data set are essentially abrupt zoom-ins or zoom-outs, which raises the question of how such JCs compare to regular, gradual zoom-ins/zoom-outs in terms of meaning and form—a question that I, however, will not address in this paper.

Finally, a note on the notation conventions adopted in this paper. In the examples I discuss, I provide a transcription of the example, followed by the number of the source video and the name of the clip from the accompanying folder in parentheses. I transcribe JC instances as <JC>, with an additional optional suffix indicating if it’s an ostensibly zooming-in JC (<JC-in>), an ostensibly zooming-out JC (<JC-out>), or a JC where the subject ostensibly shifts to the side (<JC-side> and, if applicable, <JC-back>). Some instances were not easy to categorize in these terms, so I left them

as just <JC>. Also, when relevant, I indicate cuts from or to a picture or a video clip insert (e.g., <cut-to-clip>); such inserts can be independent of JCs, but they can also be used instead of an opening or closing JC in an enclosing JC pair. Finally, when relevant, I transcribe gestures and other demonstrations in all caps.

3 Functions of the jump cut

In this section, I discuss what I found to be the most prominent semantic/pragmatic functions of the JC based on my data set. These functions are not mutually exclusive, and some of them are related to one another; so a single JC instance can perform multiple functions. Note also that this is not a fully exhaustive list, and some cases remain hard to classify, but the functions listed in this section cover the vast majority of JC instances in my data set.

3.1 Marking transition between discourse units

The broadest function of the JC is to mark transition between two discourse units of various sizes or to separate a single discourse unit from the rest of the discourse by enclosing it into a pair of JCs. Oftentimes, this function applies alongside the more specific functions discussed below, but sometimes, this seems to be the only function a given JC instance is performing.¹ For instance, a JC can mark a narrative transition:

- (2) And I was... I was too gay to laugh. <JC> A few minutes later, t.A.T.u. performed... ((1b-ii), 'later')

3.2 Marking supplements

A special case of JCs separating out discourse units are JCs marking supplements, i.e., parentheticals (as noted in Olson 2017), appositives, and sentence-level adverbials. In this use, they function similarly to (and co-occur with) the so-called “comma intonation”, to use the term from Potts 2005. For instance, in (3), a pair of enclosing JCs separates out a parenthetical; in (4), the JCs separate out an appositive;² in (5), the JC pair separates out a sentence-level *however*, which is also an instance of contrast-marking JCs (see subsection 3.7).

¹ Of course, in the case of large discourse units, JCs are also typically introduced due to practical considerations, because such large discourse units are often recorded with breaks in between.

² Here, the closing JC also serves as a separator between this discourse unit and the next one. In speech, sentence-medial supplements are enclosed by the “comma intonation” on both sides, while sentence-final supplements end with whatever the appropriate utterance-final boundary contour is in this case. Somewhat similarly, when a pair of JCs surrounds a piece of audiovisual content utterance-medially, the video will normally cut back to the initial framing and subject position after the closing JC. But when JCs separate out an utterance-final piece of content, the closing JC

- (3) A few examples are hermit crabs, pubic lice <JC-in>—the last kind of crab you ever wanna come across—<JC-out> and horseshoe crabs. ((3b-i), ‘lice’)
- (4) And as a bonus, it doesn’t produce toxic waste, <JC-in> that may be killing you and everything around you. <JC-out> But see, when we say “ceramic”... ((3a-i), ‘waste’)
- (5) And for some, it may have been a moment of queer sexual awakening. <JC-in> However, <JC-out> a year before the performance, Britney had ended her long-term relationship with Justin Timberlake. ((1b-ii), ‘however’)

3.3 Marking list items

Another case of JCs functioning similarly to and in synergy with phrasal spoken prosody is JC sequences used to separate list items, as illustrated in (6). Note, however, that in this specific case, the first JC comes before the first list item (see also subsection 3.11), while “list intonation” in spoken prosody is a right edge phenomenon that occurs between two list items.

- (6) It’s generally just coloration that an animal has to tell the world that they are <JC-in> toxic, <JC> disgusting, <JC-in> or just won’t provide any benefit should one choose to indulge. ((3b-ii), ‘toxic’)

3.4 Marking irony

As noted in Olson 2017, JCs can mark irony. They can do so at both clausal and sub-clausal level:

- (7) But they are insights that today’s <JC-in> youths <JC-out> apparently haven’t heard before. I guess because <JC-in> not enough of them are alcoholics. ((1a-i); ‘youths’)

Such JCs are similar to (and often co-occur with) irony-signaling pauses around the target material and changes in voice quality, tempo, etc. over the target material. Note that irony-marking JCs are particularly likely to shift the framing to a much closer/further one (as also noted in Olson 2017)—more drastically so than some of the more “neutral” zooming-in or zooming-out JCs—although more subtle irony-marking JCs are also quite common. It seems that the more drastic vs. subtle frame shifts are akin, respectively, to the more exaggerated vs. subtle prosodic irony marking. This is so in (7), where *youths* is produced with a more ostensibly ironic, even sarcastic prosody, and accordingly, the JC-enclosed frame shift is much more drastic, while *not enough of them are*

will typically cut to a new framing and/or subject position. In the examples discussed in this paper, I normally don’t include the closing JC for such utterance-final cases. However, I included it in (4), to highlight the contrast with (3).

alcoholics is delivered in a more deadpan fashion, and the frame shift is much less drastic. It is also possible that sub-clausal irony marking might tend to be more ostensible across the board.

3.5 *Marking demonstrations*

JCs can mark demonstrations, in the broadest understanding of this notion from Davidson 2015. This includes demonstrations introduced by overt attitude predicates or items like *like*; role shifts that are only marked prosodically and/or visually; partial quotations (in which case the JC may co-occur with gestural and/or prosodic air-quotes); demonstrations that co-occur and interact with spoken material (“co-speech”); compositionally integrated demonstrations that have their own time slot (“pro-speech”); demonstrations that are standalone discourse units (e.g., commenting on the preceding utterance, i.e., “post-speech”), etc. To give just a few examples: in (8), the demonstration enclosed in JCs is an ironic partial quotation with an iconic gestural and prosodic component; in (9), the JC separates out a compositionally integrated cringing facial expression that serves as a predicate roughly meaning ‘such that it warrants the following reaction: DEMONSTRATION’ (see Esipova 2022 for more examples of such reaction-based compositional integration of demonstrations); in (10), the JC separates out a “post-speech” snicker (cf. utterance-final face emoji, as discussed, e.g., in Grosz et al.).

- (8) And I realize that to most people, complaining about <JC-in> [being cancelled WAAH]^{WEEP-GESTURE+VOICE} <JC-out>, it sounds incredibly whiny and self-absorbed. ((1a-ii), ‘cancelled’)
- (9) There was a time, not so long ago, when the UK’s trans representation in the media was <JC-in> CRINGE-FACE. ((1b-i), ‘cringe’)
- (10) So it’s a good thing my parents spent all that money. <JC-in> SNICKER+COLOR-CHANGE. ((2b-i), ‘money’)

3.6 *Marking contrast*

JCs can mark contrast. For instance, they can be used to introduce adversative clauses, either enclosing the adversative connective, as we have already seen in (5), or simply separating the adversative clause itself from the preceding (and possibly following) material, as in (11).

- (11) And I think, maybe, in this instance, what’s happening is that Kerrigan knew how to write a Northern family, when they’re all cis, <JC-in> but he doesn’t know trans people. ((1b-i), ‘cis’)

JCs can also mark contrast in the absence of adversative connectives (note also the speaker’s contrast-marking gestures that interact with the on-screen text):

(12) But see, when we say “ceramic”, and when <JC-in> they say “ceramic”, we’re not actually talking about the same thing. ((3a-i), ‘ceramic’)

3.7 *Marking intensification*

JCs can be used to mark degree intensification, often alongside prosodic and/or gestural degree intensification (see, e.g., Esipova2019a,b; 2022), as in (13). Note also the connection to demonstration-marking JCs here, as expressive degree intensification (via spoken expressives, prosody, and facial expressions) can be analyzed in a demonstration-based way (Esipova 2022).

(13) Now, this craze has gone <JC-in> so far beyond just the celebrity class. ((3a-ii), ‘beyond’)

3.8 *Marking emphasis*

The notion of emphasis is somewhat nebulous and presumably plays a role in many other functions of the JC. That said, this is the best characterization I have for now for examples like (14).

(14) ...because I am <JC-in> not a gamer. ((2a-ii), ‘gamer’)

3.9 *“Ramp-up sequences”*

One can also have a sequence of zooming-in or zooming-out JCs—which can be more dramatic than a regular, gradual zoom-in/zoom-out—for instance, to convey ramping up emotions:

(15) Celebrities are under attack. <JC-in> This is the new Salem. <JC-in> This is Orwell’s nightmare. ((1a-ii), ‘attack’)

3.10 *Marking disfluencies*

While, as noted above, JCs can be used to mask disfluencies when editing, a fairly common meta use of a JC is to instead draw attention to a disfluency, such as struggling with pronouncing a word, as a self-deprecating joke, for instance:

(16) In the early days, PTFE was manufactured by using <JC-in> perfluoro-octa-noic acid, PFOA. ((3a-i), ‘pfoa’)

3.11 *Expressing identity and/or individual style*

While the widespread use of JCs on YouTube is undeniably due to its practicality and semantic versatility, its proliferation has now arguably led to YouTubers also using JCs because other YouTubers do. There also appear to be genre effects; for instance, JCs seem to occur much more densely in commentary videos than in video essays, likely because the commentary genre often uses more dynamic and less polished editing overall, has higher joke density, etc. Finally, the use of JCs can be part of one’s personal style. For example, Lindsay Nikole uses far more JCs than the

other creators on my list. Many individual JCs in her videos don't necessarily have a very specific semantic/pragmatic function, but instead align with prosodic events and/or syntactic boundaries, not unlike beat gestures, resulting in a unique, dynamic editing style that matches the choppy prosody of her delivery.

3.12 Summary

When it comes to the range of functions that JCs can perform, perhaps the closest counterpart from spoken communication would be simply a pause, as pauses emerge in all cases above, too (see, e.g., Beltrama & Hanink 2019; Esipova 2019a,b; Harris 2021). However, JCs often occur without ostensible pauses (as shown in subsection 4.2 below), so the two are not fully homologous.

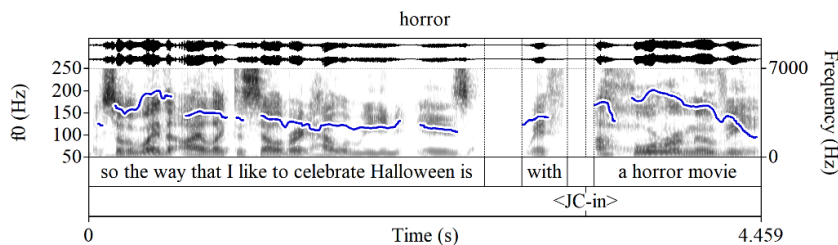
4 Prosodic integration of sub-ip jump cuts

Many JCs occur at intonational phrase (IP, largest prosodic unit, typically mapping to clause boundaries in the syntactic structure) or at least intermediate phrase (ip, smaller prosodic unit within IPs) prosodic boundaries (see the ToBI guidelines in Beckman & Ayers 1997 for details): e.g., (11) and (6), respectively. However, there are plenty of cases of sub-ip JCs. There are two major strategies of integrating such sub-ip JCs with the speech stream: (i) co-occurring with and supporting acoustic prosodic discontinuity; and (ii) anchoring to a pitch accent, similarly to beat gestures. I elaborate on both strategies below, and I also briefly discuss how JCs can use either strategy when integrating with expressive beat sequences.

4.1 Aligning with prosodic discontinuity

Sub-ip JCs can align with prosodic discontinuity within an ip, in particular, pauses, with lengthening of the preceding segmental material, a preceding intake of breath, etc. (essentially, prosodic boundaries that would be labeled with break index 2 in ToBI). JCs thus create a synergistic visual discontinuity matching the acoustic discontinuity. For instance, in (17), the JC aligns with a pause and is preceded by ostensible segment lengthening on *with*:

(17)



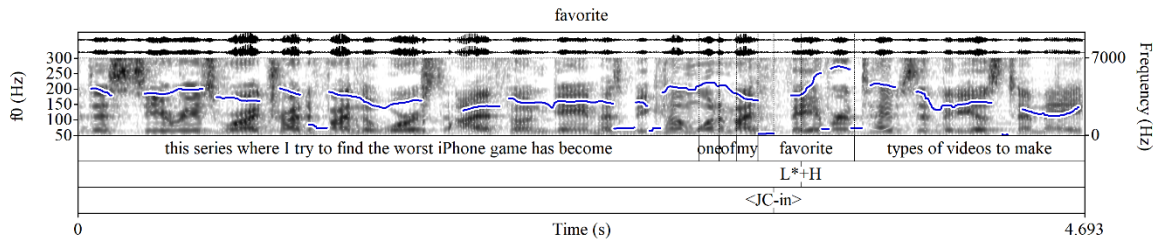
((2b-ii), 'horror')

Note that such prosodic discontinuity is typically necessary to integrate pro-speech demonstrations with speech (see Esipova 2019a; Harris 2021), so you often see it in examples like (8) or (9).

4.2 Anchoring to a pitch accent

Sub-ip JCs can also anchor to a pitch accent, and when they do, they will typically slightly precede one, as in (18). This is reminiscent of prosodic integration of beat gestures, which also often anchor to pitch accents in a similar way (see, e.g., Loehr 2004 and references therein); note that (18) does contain a beat gesture anchoring to the pitch accent on *favorite*, as well.

(18)

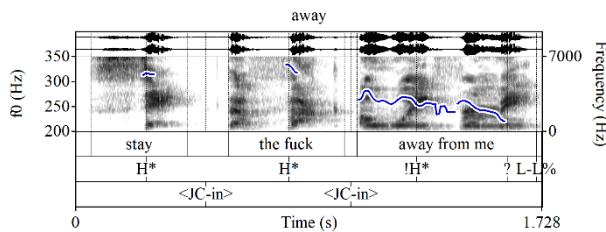


((2a-ii), ‘favorite’)

4.3 Integration with expressive beat sequences

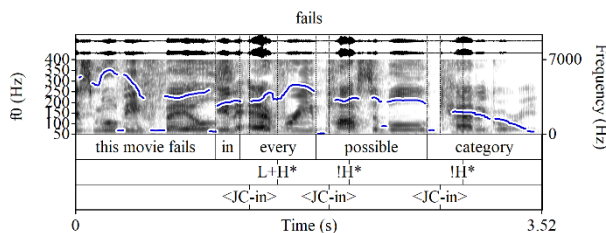
Ramp-up JC sequences (discussed in subsection 3.9) can co-occur with what Esipova (2022) calls “expressive beats” (in prosody, gesture, and written text), creating a choppy visual rhythm to match the rhythm created by the spoken prosody (and sometimes gesture). When JCs integrate with such expressive beat sequences, they can use either of the strategies above. I.e., they can align with the pauses between the beat units (similarly, to how punctuation marks or emoji occur between units in written expressive beat sequences), as illustrated in (19). Or they can anchor to the pitch accents of the beat units (similarly to gestures), as illustrated in (20).

(19)



((3b-ii), ‘away’)

(20)



((2a-i), 'fails')

5 Conclusion

As YouTubers are often more limited in their technical resources than conventional filmmakers, while at the same time less constrained by the pressure to follow the traditional rules of editing, they have been extensively using the jump cut as a replacement for a lot of conventional editing techniques. Despite my study being very modest in scale, it does warrant a conclusion that by now, the jump cut has evolved into one of the fundamental elements of visual prosody on YouTube, helping create a visual prosodic structure that integrates with the co-occurring speech stream, both at the level of meaning and at the level of prosody. While this integration is diverse and complex, it is not random: as the many examples discussed in this paper show, it draws on existing patterns of integration of segmental, suprasegmental, and gestural material into a coherent multi-channel signal. This allows JCs (and other editing techniques) to become a synergistic part of this signal in YouTube videos. Having different realizations of the same meaning via different channels can create a strong cumulative effect in spoken communication, especially in the case of highly iconic depictions, whose aim is to create a direct sensory experience for the addressee (see, e.g., Dingemanse 2015; Clark 2016 on depiction). Adding editing events that match the different aspects of the spoken-gestural signal further enhances this sensory experience. I, thus, hope that this paper can serve as an inspiration for linguists to explore more cases of such speech-gesture-editing integration in the future.

Source videos

(1) Video essays

- a. 'ContraPoints' YouTube channel:
 - i. 'Jordan Peterson | ContraPoints' (2018); <https://youtu.be/4LqZdkkBDas>
 - ii. 'J.K. Rowling | ContraPoints' (2020); <https://youtu.be/7gDKbTl2us>
- b. 'verilybitchie' YouTube channel:
 - i. 'Good LGBT Representation is Boring (and why that's a problem)' (2021); <https://youtu.be/cR3b2Gblq0>
 - ii. '00s Bisexual Chic' (2022); <https://youtu.be/hqdD5d3iRoU>

(2) Commentary

- a. 'Danny Gonzalez' YouTube channel:

- i. ‘An Absolutely Terrifying Low Budget RomCom’ (2022); <https://youtu.be/uubMLkM5L9E>
 - ii. ‘Trying To Find The Worst iPhone Game 3’ (2023); <https://youtu.be/m8LBvGX2hCQ>
 - b. ‘Kurtis Conner’ YouTube channel:
 - i. ‘A Deep Dive Into Disney Adults’ (2021); <https://youtu.be/BvNmLwOLz3w>
 - ii. ‘This Low-Budget Horror Movie is Terrifying...For All The Wrong Reasons’ (2021); <https://youtu.be/yvZP3YQNuRo>
- (3) Edutainment
- a. ‘Future Proof’ YouTube channel:
 - i. ‘The TRUTH about Ceramic Cookware’ (2022); <https://youtu.be/TeXObJa4D4k>
 - ii. ‘Why Are People OBSESSED with Diet Coke?’ (2023), <https://youtu.be/L3meiJwNLtQ>
 - b. ‘Lindsay Nikole’ YouTube channel:
 - i. ‘Is CRAB the final form?’ (2023), <https://youtu.be/pv--L0FyIu4>
 - ii. ‘Zoologist Answers: WTF is THAT?? (& Don’t Touch Them!)’ (2023), <https://youtu.be/i6V-RBjecpI>

References

- Beckman, Mary E. & Gayle Ayers. 1997. Guidelines for ToBI labelling. Version 3.0. The Ohio State University Research Foundation.
- Beltrama, Andrea & Emily A Hanink. 2019. Marking imprecision, conveying surprise: *Like* between hedging and mirativity. *Journal of Linguistics* 55(1). 1–34. <https://doi.org/10.1017/S0022226718000385>
- Clark, Herbert H. 2016. Depicting as a method of communication. *Psychological review* 123(3). 324–347. <https://doi.org/10.1037/rev0000026>
- Dingemanse, Mark. 2015. Ideophones and reduplication: Depiction, description, and the interpretation of repeated talk in discourse. *Studies in Language*. 39(4). 946–970. <https://doi.org/10.1075/sl.39.4.05din>
- Davidson, Kathryn. 2015. Quotation, demonstration, and iconicity. *Linguistics and Philosophy* 38(6). 477–520. <https://doi.org/10.1007/s10988-015-9180-1>
- Davidson, Kathryn. 2022. Depictive versus patterned iconicity and dual semantic representations. Talk given at The 96th Annual Meeting of the LSA, Washington, DC.

- Esipova, Maria. 2019a. *Composition and projection in speech and gesture*. PhD thesis, New York University. <https://ling.auf.net/lingbuzz/004676>
- Esipova, Maria. 2019b. Towards a uniform super-linguistic theory of projection. In Julian J. Schlöder, Dean McHugh & Floris Roelofsen (eds.), *Proceedings of the 22nd Amsterdam Colloquium*, 553–562.
- Esipova, Maria. 2022. Composure and composition. Ms. <https://ling.auf.net/lingbuzz/005003>
- Grosz, Patrick Georg, Gabriel Greenberg, Christian De Leon & Elsi Kaiser. 2023. A semantics of face emoji in discourse. *Linguistics and Philosophy* 1–53. <https://doi.org/10.1007/s10988-022-09369-8>
- Harris, Alexis. 2021. *Unspoken intonation: The prosody of pro-speech gesture*. Senior thesis, Princeton University.
- Loehr, Daniel P. 2004. *Gesture and intonation*. PhD thesis, Georgetown University.
- Olson, Dan. 2017. Why The Jump Cut Is Here To Stay. Video essay, YouTube channel ‘Folding Ideas’. <https://youtu.be/XvK8xtVbopA>
- Patel-Grosz, Pritty, Salvador Mascarenhas, Emmanuel Chemla & Philippe Schlenker. 2022. Super Linguistics: an introduction. *Linguistics and Philosophy, special issue on Super Linguistics*. <https://doi.org/10.1007/s10988-022-09377-8>
- Potts, Christopher. 2005. *The logic of conventional implicatures*. Oxford: Oxford University Press.