Language Under Discussion
http://www.ludjournal.org ISSN: 2329-583x

On the fringes of grammar: The role of prosody in affective communication

Nicolás Rivera Bobadilla[a]

[a] Universitat Pompeu Fabra, nicolas.rivera@upf.edu.

DOI: 10.31885/lud.7.1.267

Paper received: 8 July 2025
Published online: 18 November 2025

Abstract. This article supports Wiltschko's argument, extending her proposal to affective prosody. This is needed, since Wiltschko's proposal neglects prosody as a mode of emotional expression. Particularly, I analyze aspirations and vowel lengthening to show that their emotional interpretation is necessarily context-dependent, as they express diametrically opposed emotions in various linguistic contexts. Thus, as there is no correspondence between distributions and functions, they are expected to fall outside grammar. Moreover, the use of another prosodic cue—rising intonation—is also discussed to illustrate how it performs a grammatical function, as it encodes epistemic states, hence being different from emotional aspirations and vowel length.

Keywords: prosody, Universal Spine, interaction, emotions, aspirations, vowel length, intonation

1. Introduction

In her focus article, Martina Wiltschko (2024) proposes that emotions are constructed by the same computational system responsible for the construction of utterances, namely, the Universal Spine: a hierarchically organized syntactic structure built for the study of grammatical categories across non-related languages (Wiltschko, 2014). To support her claim, Wiltschko focuses on describing how verbal constructions express emotions, showing that there are no dedicated grammatical categories for them. Thus, she convincingly demonstrates that emotions are not part of the architecture in charge (of the communication) of thought. To my knowledge, this is the first attempt at a grammatical theory—built from observations in typologically distant languages—to explain the relation between language and emotions.

Language Under Discussion, Vol. 7, Issue 1 (August 2024), pp. 111–119
Published by the Language Under Discussion Society

In Wiltschko's novel proposal, however, prosody receives only limited attention, despite its widely recognized importance as a crucial mode of emotional expression (Bestelmeyer et al., 2017; Larrouy-Maestri et al., 2024). In order to remedy this empirical limitation and to support Wiltschko's argument, the present note examines some uses of prosodic cues in face-to-face interaction. I will focus on two prosodic means of expression traditionally associated with emotional expression: aspirations and vowel lengthening. I aim to show that the emotional interpretation of such resources does not correlate with distributional patterns and is necessarily context-dependent. Thus such an interpretation cannot be associated with grammar. Moreover, I will present cases in which prosody can perform a grammatical function, and that such a function corresponds to non-emotional meaning.

The approach I take here is informed by Conversation Analysis and Interactional Linguistics (Couper-Kuhlen, 2009; Hepburn & Bolden, 2013; Couper-Kuhlen & Selting, 2018; Couper-Kuhlen et al., 2024; henceforth CA/IL), since the role of prosody in grammar has been mostly treated in these disciplines and scarcely explored in Formal Linguistics (cf. Hedberg & Sosa, 2008; Truckenbrodt, 2013; Goodhue, 2018). Moreover, an CA/IL approach will allow us to observe the mismatches between the extended Universal Spine (i.e., the grammar of propositional and interactional language) and the affective uses of prosodic resources. For this purpose, the data used in this note is composed of excerpts from naturally occurring conversations, with transcriptions that reflect prosodic information.

This note is structured as follows: in the next Section, I present an overview on the relation between grammar and interaction. In Section 3, I present the different uses of two prosodic features—aspirations and vowel length—in several conversation excerpts, showing that their interpretation is contextually determined. In Section 4, I show how prosody can perform grammatical functions according to the Universal Spine Hypothesis. I will conclude, in Section 5, with some remarks about the formalization of these resources.

2. How do we determine if an interactional resource is part of or falls outside grammar?

In describing the systematicity that underlies the use of linguistic forms exclusively dedicated to managing the communication process, CA/IL has shown that speakers make use of various resources in face-to-face interaction to communicate their affective stance: verbal (morphosyntactic and lexical elements used in a context-sensitive fashion, at a particular place within an action), prosodic (prosody and phonetics that are not distinct for single affects and are tied to the sequential environment), and non-vocal (posture, gestures, facial expressions, gaze, and head movement) (Peräkylä & Sorjonen, 2012, pp. 5–9). As acknowledged by Wiltschko (2024) too, these means of expression are also used in social interaction on a regular basis for functions different from affective displays. Therefore, when speakers express emotions, contextual information is crucial for interpreting their affective stance.

Despite the context-dependence of many of these resources, the bottom-up insights from CA/IL about their use have expanded the empirical horizon of various grammatical theories: constructionist (Steensig et al., 2025), functionalist (Evans et al., 2018; Heine, 2023), and formalist (Wiltschko, 2021). The Interactional Spine Hypothesis (Wiltschko, 2021), an extension of the Universal Spine (Wiltschko, 2014), is one of the grammatical theories that has benefited from this approach in particular. In order to distinguish whether or not a form used in social interaction fits into grammar, it is necessary to observe the contexts in which such a form is used for

finding recurrent patterns for the accomplishment of a given function (Couper-Kuhlen, 2012, p. 2).

What fits into grammar (or is in the process of grammaticalization) is a matter of great debate (for an overview on grammaticalization, see Heine & Narrog, 2012) and depends to a considerable extent on the particular grammatical theory one uses. In this case, I will focus on Wiltschko's grammatical theory and on what prosodic cues can be associated with it. In order to identify which of these resources can be candidates for a grammatical description, they need to display a function that exhibits a core meaning across contexts of use in a defined distribution (Wiltschko et al., 2018). In this sense, it will be possible to describe a linguistic form in terms of grammar if its use in a determined distribution entails an interactional function that cannot be performed by it otherwise and does not change its core effect across contexts of use.

In the next section, I turn to observe how aspirations and vowel lengthening—two prosodic resources associated with emotional expression—have context-dependent meaning, thus falling outside grammar.

3. Aspirations and vowel lengthening: context dependent interpretations

As mentioned above, prosody is a crucial resource for affective communication (Freese & May nard, 1998). Aspirations and vowel lengthening are used in widely different contexts and distributions, and they can express widely different affective content depending on their context of use (Waaramaa et al., 2010; Hoey, 2014). These two prosodic resources are highly context-dependent, thereby differing from grammaticalized forms of expression, which relate with distinctive contexts of use and meaning (Wiltschko et al., 2018).

To illustrate this point, consider the excerpt in (1). Here, one of the interactants shares with the other a traumatic experience: J's house burned down during the night. Both participants orient themselves to express their emotions throughout the conversation, which is done through various prosodic phenomena: turn-overlapping, vowel lengthening, aspirations, breathy pronunciation, and broken words. Such prosodic features, which appear at specific moments of the interaction, enable the interpretation of the interactants' affective stances: J grieving about her lost home and P aiming to empathize with her.

(1) Jefferson (1988, p. 428; in Lindström & Sorjonen 2013, p. 352)[1]:

```
1  J: It happened within: minutes. .hhh Within a half hour
2     the house was go:ne I guess,=
3  P: =Ohh:: Ghod,
4  J: So it's just l[i:ke, we wouldn't we just wouldn't've been=
5  P:               [.hhh
6  J: =here .hh You kno:w,
7  P: .t! [OH::::::::: ba:by.              ]
8  J:     [There's no way it was it was jus]:t,
9     we're just lucky I guess:,
```

[1] The transcriptions of the conversations follow Jeffersonian conventions (Hepburn & Bolden, 2013). For the purposes of this note, I highlight here the following: h aspiration; .h inhalation; [ ] overlapping talk; = beginning of a new turn immediately after the previous one without a gap in between; (.) brief silence interval; . final falling intonation; , slightly rising intonation; :: vowel lengthening; ↑ rise in pitch; underlined prosodic stress; CAPITALS words louder than the surrounding utterance.

```
10  P: .hhhh Okay wait [a minute I don't know if you're cryi-ing
11  J:                 [So,
12  P: but I hhh(h)a[hhhm uh hm: .hh=
13  J:              [(hhhhhh hum)
14  J: =.h I was guh- I- middle of the night last la-ast night I
15     wantehhd hhto hhc(h)all (h)y(h)ou .mhhh!
```

Aspirations—represented with `h` and abrupt spurts with `(h)`—are used inside words in widely different linguistic contexts: within a noun ("`Ghod`" in line 3), a verbal phrase ("`wantehhd hhto hhc(h)all (h)y(h)ou`" in line 15), or interjections ("`ahhhm uh hm`" in line 12). Moreover, it can also appear as a stand-alone sound, represented with `hhh`. In this excerpt, this prosodic cue is used by J to express sorrow, and by P to show empathy towards her. Vowel lengthening—represented with `:`—is also used within various parts of speech, and it functions either as a means of hesitation due to a turn overlapping (as "`so it's just li:ke`" in line 4), or to intensify the negative affective stance conveyed in the turn (as "`OH::::::::: ba:by`" in line 7).

In other conversational contexts, these two resources are used for expressing positive emotions. Such is the case in (2) and (3). In these excerpts, interactants display an invitation to laugh through the use of aspirations and vowel length, thereby co-constructing joy and amusement.

(2) Jefferson (1979, p. 83):

```
1  K: But going sou::th.
2  F: Et the end of Harbor Bouleva(HH)rd=
3  K: =↑ahh ha::ha::ha::
```

(3) Jefferson (1979, p. 83):

```
1  B: Dju watch by any chance Miss International Showcase las'night?
2  E: N:no I didn'[I wz reading my-
3  B:             [You missed a really great Pro(H)=
4  E: =↑O(hh)h i(h)t wa(hh)s?=
5  E: =ehh heh heh heh!
```

In these excerpts, aspirations within words are interpreted as signaling a "candidate laughable utterance" (Jefferson, 1979, p. 83). Such is the case in (2), "`Et the end of Harbor Bouleva(HH)rd`", and (3), "`You missed a really great pro(H)[gram]`". These two utterances are followed by the recipient's laughter, thereby being successful invitations to laugh.

In addition to aspirations, vowel lengthening can also be used to invite participants to laugh. Such is the case in (4), where the interactants describe the symbol of Sagittarius, and in doing so they invite each other to make fun of it. Both participants compete to make a joke that will amuse the other, without success.

(4) Jefferson (1979, p. 88):

```
1  Mike: Wud is dat.
2              (0.9)
3  Vic:  Uh:::: ih-ah-ih guy widda bow'n arrow Enna half en a:sss. .hh A
```

```
 4        half en a:ss o[f a ho::]ssss.
 5  Mike:                [is that-]=
 6  Vic:   =he[h -he:h (h)uh  w(h)uddevuh   ]
 7  Mike:     [↑Yeh but wuddih they call it ]=
 8  Vic:      =duh hell 'e i(h)::s,]
 9  Mike: what izzit the ra:m?]=
10  Vic:   =[.hhh!
11  Mike:   [Saggitarius's uh ra:m?
12        (·)
13  Vic:   No. Ra:m is, ram is uh:: a ram.Yihknow wid ho:rns.
```

Crucially, vowel lengthening observed in this excerpt can be interpreted either as an indication of doubt (as in "uh:::" in lines 3 and 13) or as a way of stressing a comical element i.e., the animal possibly depicted in the symbol of Sagittarius) to make the other person laugh. This latter use of vowel lengthening reflects the speaker's aim to emphasize a critical point in the conversation, namely, naming the funny animal and making the other person laugh.

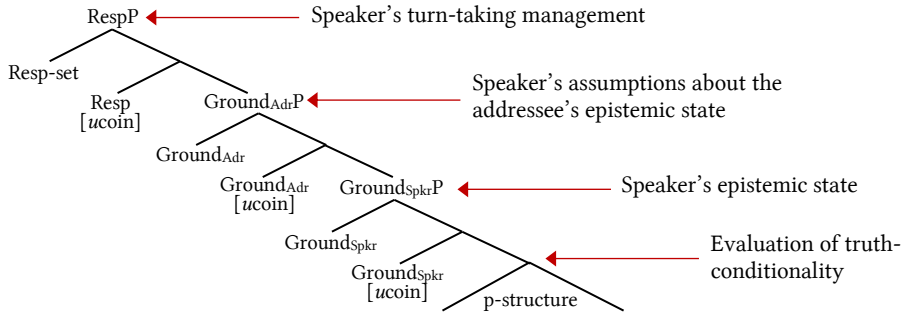
We have seen that the emotional interpretation of aspirations and lengthened vowels is necessarily tied to the context of use, as evidenced by their different affective meaning. Hence, it is expected that these prosodic cues will not have a corresponding grammatical function for such a content. This reading is in line with Wiltschko (2024), who established that the emotional interpretation of lexical items and syntactic constructions falls outside grammar. In this sense, the observations about emotional prosody pointed out here expand Wiltschko's claim to another mode of communication.

There are, however, prosodic cues that are mapped onto grammar (Heim, 2019; Wiltschko, 2022). This is observed in cases where certain intonation patterns, with a fixed distribution, are used for expressing epistemic states or for managing turn-taking. I now turn to examine such cases from the perspective of the Spine (Wiltschko, 2014, 2021).

4. Prosody can enter grammar

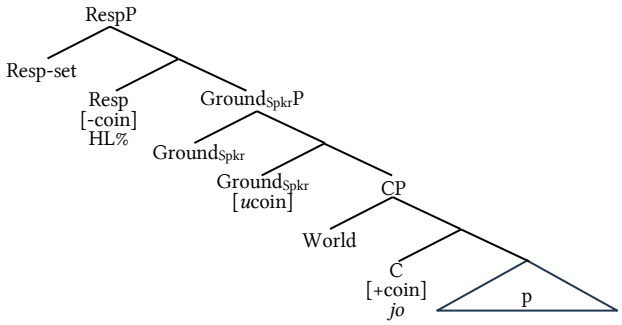
As stated above, the Interactional Spine Hypothesis is a grammatical framework for the description of linguistic forms dedicated to managing communication. The Spine is a hierarchically organized architecture designed for the formalization of Units of Language that work in regulating turn-taking and common ground between interactants (Wiltschko, 2021). It is composed of three articulated layers which encode different interactional functions with respect to a given proposition (p-structure) that is shared in interaction: $\text{Ground}_{\text{Spk}}$, which encodes the speaker's beliefs towards such a proposition, $\text{Ground}_{\text{Adr}}$, which encodes the speaker's assumptions about their addressee's beliefs, and Response, which regulates turn-taking. These functions are syntactically interpreted through a Unit of Language that assigns a value to an unvalued coincidence feature [*u*coin] via agreement (for a more detailed explanation of this process, see Wiltschko, 2014). The value of the coincidence feature serves to relate the complementizer with its specifier. By extending syntax beyond a propositional structure, this architecture allows to describe linguistic forms that are commonly thought to fall outside syntax, but have shown to have a systematic use in CA/IL research: discourse markers, modal particles, interjections, response markers, confirmationals, intonation contours, etc. The main functions are illustrated in (5).

(5) The Interactional Spine (Wiltschko, 2021, p. 151, 202)

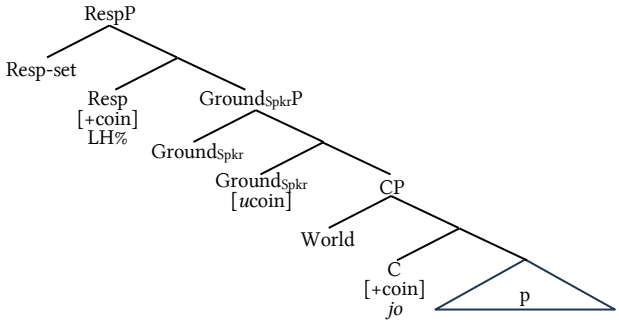


As part of the architecture of the Spine, intonation has been described as a meaningful mechanism for turn-taking (Wiltschko, 2021, 2024, p. 41), and commitment and engagement (Wiltschko & Heim, 2016; Heim, 2019). Thus, intonation has been shown to be grammatically significant when used in specific contexts, as it acts as a Unit of Language that assigns a value to a grammatical function. This is observed in, for instance, the response marker *jo* in Austrian German, when it is used with a falling intonation (in 6) and with a rising intonation (in 7):

(6) Falling *jo* (Wiltschko, 2021, p. 171)



(7) Rising *jo* (Wiltschko, 2021, p. 172)

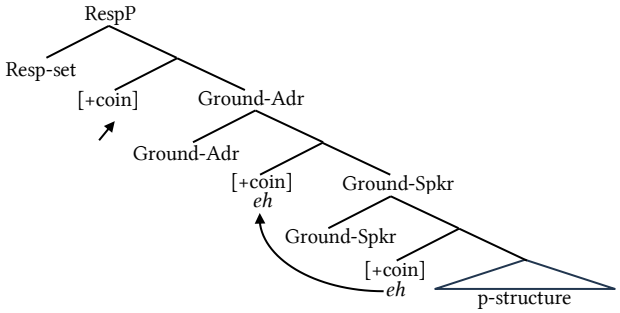


The response marker *jo* is uttered with falling intonation in (6), which is interpreted as the speaker assessing the information they just received as something already known. By contrast, the rising intonation used with *jo* in (7) is interpreted as a request for further information, since the speaker has received some unexpected news.[2] In this way, when rising intonation is used in a response marker, such as *jo*, it expresses grammatically significant content.

[2] Note that the grammatical function proposed for rising intonation is compatible with an emotional interpretation shaped by context. Unexpected information can be evaluated as either positive or negative (e.g., a friend getting a new dog is positively surprising, whereas a friend being involved in a car accident is negatively surprising).

Another use of rising intonation that has been described as grammatically relevant is when it is used at the end of assertions. In such a context, this intonation can serve to manage epistemic attitudes in interaction (Heim, 2019). Particularly, when rising intonation is realized through particles, such as *eh* or *huh* in English, it can be interpreted in two ways: (1) as a request for the addressee to confirm the truth value of the host utterance, or (2) as a request for the addressee to confirm their awareness that the content of the host utterance is true (Wiltschko & Heim, 2016). These two interpretations are mapped as part of the Interactional Spine (Wiltschko 2021, 2022), as is illustrated in (8).

(8) *eh?* as a confirmational (Wiltschko, 2022, p. 294)



As we have seen, rising intonation and its pragmatic effects can be described from a grammatical point of view. Moreover, the cases mentioned in this section are not dedicated to the expression of emotion, but rather to the management of epistemic (confirmation of belief) and deontic (request for an answer) stances between interactants. Even though an affective interpretation of these intonations is possible—for example, one might show enthusiasm by responding with a rising *jo!,* or can express happiness for somebody by asking "*You had a great performance today, eh?*"—the syntactic analyses in (7) and (8) propose that such an interpretation is built from the interplay between context and the core functions of the two types of rising intonation described here.

Unlike emotional aspirations and vowel lengthening, which can be used in any position within the utterance for expressing different affective stances, the two interpretations of rising intonation come from determined contexts of use: as part of a response marker or at the end of an assertion, thus performing a grammatical function (for a detailed discussion on this regard, see Wiltschko & Heim, 2016; Heim, 2019, Wiltschko, 2022). Therefore, the grammaticalized use of prosody sharply contrasts with the prosodic cues that are interpreted as emotional forms of expression.

5. Concluding remarks: To what extent is it possible to formalize all these resources?

This note aimed to support Wiltschko's (2024) argument that emotions do not enter grammar by addressing an empirical limitation of the target article, namely, the role of prosody in emotional expression. I have shown that the affective interpretation of aspirations and vowel length necessarily depends on the conversational context. Therefore, their emotional function cannot be captured within a grammatical architecture. Hence, and as Wiltschko (2024) proposed, affective interpretations of prosody are not part of grammar, since they are necessarily context dependent.

Unlike in the case of grammaticalized rising intonation—which is used in response markers and at the end of assertions—emotional aspirations and vowel lengthening are used in various positions within an utterance, without observable constraints in this regard. Moreover, the affective display of these two prosodic cues is not influenced by the positions in which they are used. Rather, it appears to depend on the speaker's communicative aim and their interlocutor's interpretation within the particular conversational context.

The multidimensionality of emotional experience coincides with the multidimensionality of language, which in turn should coincide with any attempt to formalize this relationship. From this perspective, to distinguish the linguistic resources that are part of grammar from the ones that rely entirely on context composes an analytical challenge. In this sense, an architecture, such as the one proposed in Wiltschko's focus article, can help shed light on this matter, as it allows us to distinguish which resources fall within grammar and which fall outside it. In line with Wiltschko's (2024) proposal, the data presented in this brief note indicate that emotional prosody belongs to the latter group.

Acknowledgements

I would like to express my gratitude to the anonymous reviewers who helped me improve early versions of this manuscript. I am also grateful to the Editorial Team of *Language Under Discussion* for making these discussion notes possible. The work behind this note is supported by the Chilean National Agency for Research and Development (ANID), Becas Chile 72220098.

References

Bestelmeyer, P., Kotz, S., Belin, P. (2017). Effects of emotional valence and arousal on the voice perception network. *Social Cognitive and Affective Neuroscience*, *12*(8), 1351–1358.

Couper-Kuhlen, E. (2009). A sequential approach to affect: The case of 'disappointment'. In M. Haakana, M. Laakso, J. Lindström (Eds.), *Talk in interaction: Comparative dimensions*, (pp. 94–123). Helsinki: Finnish Literature Society.

Couper-Kuhlen, E. (2012). Grammaticalization and conversation. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Grammaticalization* (pp. 424–437). Oxford: Oxford University Press.

Couper-Kuhlen, E., & Selting, M. (2018). *Interactional Linguistics: Studying Language in Social Interaction.* Cambridge University Press.

Couper-Kuhlen, E., Fox, B. A., Raymond, C. W., Sorjonen, M-L. & Thompson, S. A. (2024). Analyzing grammar in social interaction. In J. D. Robinson, R. Clift, K. H. Kendrick & C. W. Raymond (Eds.), *The Cambridge Handbook of Methods in Conversation Analysis* (pp. 641–671). Cambridge: Cambridge University Press.

Evans, N., Bergqvist, H., & San Roque, L. (2018). The grammar of engagement I: Framework and initial exemplification. *Language and Cognition*, *10*(1), 110–140.

Freese, J., & Maynard, D. W. (1998). Prosodic features of bad news and good news in conversation. *Language in Society 27*(2), 195–219.

Goodhue, D. (2018). *On Asking and Answering Biased Polar Questions.* Doctoral Dissertation. McGill University.

Hedberg, N., Sosa, J.M. (2008). The prosody of topic and focus in spontaneous English dialogue. In C. Lee, M. Gordon & D. Büring (Eds.), *Topic and Focus* (pp. 101–120). Dordrecht: Springer.

Heim, J. (2019). *Commitment and Engagement: The Role of Intonation in Deriving Speech Acts.* Doctoral dissertation. University of British Columbia.

Heine, B. & Narrog, H. (2012). *The Oxford Handbook of Grammaticalization.* Oxford: Oxford University Press.

Heine, B. (2023). *The Grammar of Interactives.* Oxford: Oxford University Press.

Hepburn, A. & Bolden, G. (2013). The Conversation Analytic approach to transcription. In J. Sindell & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 57–76). Chichester, UK: Wiley-Blackwell.

Hoey, E. M. (2014). Sighing in interaction: Somatic, semiotic, and social. *Research on Language and Social Interaction 47*(2), 175–200.

Jefferson, G. (1979). A technique for inviting laughter and its subsequent acceptance declination. In G. Psathas (Ed.), *Everyday Language: Studies in Ethnomethodology* (pp. 79–96). New York: John Wiley & Sons.

Jefferson, G. (1988). On the sequential organization of troubles-talk in ordinary conversation. *Social Problems 35*(4), 418–441.

Larrouy-Maestri, P., Poeppel, D., & Pell, M. D. (2024). The sound of emotional prosody: Nearly 3 decades of research and future directions. *Perspectives on Psychological Science 20*(4), 623–638.

Lindström, A., & Sorjonen, M.-L. (2013). Affiliation in conversation. In J. Stivers & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 250–369). Chichester, UK: Wiley-Blackwell.

Peräkylä, A. & Sorjonen, M.-L. (Eds.). (2012). *Emotion in Interaction.* Oxford: Oxford University Press.

Steensig, J., Jørgensen, M., Lindström, J., Mikkelsen, N., Suomalainen, K. & Sørensen, S. (2025). *Grammar in Action: Building Comprehensive Grammars of Talk-in-interaction.* Amsterdam: John Benjamins.

Truckenbrodt, H. (2013). An analysis of prosodic F-effects in interrogatives: Prosody, syntax and semantics. *Lingua*, *124*, 131–175.

Waaramaa, T., Laukkanen, A., Airas, M., & Alku, P. (2010). Perception of emotional valences and activity levels from vowel segments of continuous speech. *Journal of Voice 24*(1), 30–38.

Wiltschko, M. (2014). *The Universal Structure of Categories: Towards a Formal Typology*. Cambridge: Cambridge University Press.

Wiltschko, M. & Heim, J. (2016). The syntax of confirmationals: A neo-performative analysis. In G. Kaltemböck, E. Keizer & A. Lohmann (Eds.), *Outside the Clause* (pp. 305–340). Amsterdam: John Benjamins.

Wiltschko, M., Denis, D., & D'Arcy, A. (2018). Deconstructing variation in pragmatic function: A transdisciplinary case study. *Language in Society 47*(4), 569–599.

Wiltschko, M. (2021). *The Grammar of Interactional Language.* Cambridge: Cambridge University Press.

Wiltschko, M. (2022). What is the syntax-pragmatics interface? *Empirical Issues in Syntax and Semantics 14*(1), 273–208.

Wiltschko, M. (2024). Emotions do not enter grammar because they are constructed (by grammar). *Language Under Discussion 7*(1), 1–62.