TYPE Hypothesis and Theory
PUBLISHED 19 February 2024
DOI 10.3389/fcomm.2024.1338844

OPEN ACCESS

EDITED BY
Maria Grazia Sindoni,
University of Messina, Italy

REVIEWED BY
Kyle Jasmin,
University of London, United Kingdom
Valentin Werner,
University of Bamberg, Germany

*CORRESPONDENCE
Claudia Lehmann
✉ claudia.lehmann@uni-potsdam.de

RECEIVED 15 November 2023
ACCEPTED 01 February 2024
PUBLISHED 19 February 2024

CITATION
Lehmann C (2024) What makes a multimodal construction? Evidence for a prosodic mode in spoken English.
Front. Commun. 9:1338844.
doi: 10.3389/fcomm.2024.1338844

What makes a multimodal construction? Evidence for a prosodic mode in spoken English

Claudia Lehmann*

Chair of Present-Day English, Institute of English and American Studies, University of Potsdam, Potsdam, Germany

Traditionally, grammar deals with morphosyntax, and so does Construction Grammar. Prosody, in contrast, is deemed *paralinguistic*. Testifying to the "multimodal turn," the past decade has witnessed a rise in interest in multimodal Construction Grammar, i.e., an interest in grammatic constructions other than exclusively morphosyntactic ones. Part of the debate in this recent area of interest is the question of what defines a multimodal construction and, more specifically, which role prosody plays. This paper will show that morphosyntax and prosody are two different semiotic modes and, therefore, can combine to form a multimodal construction. To this end, studies showing the independence of prosody for meaning-making will be reviewed and a small-scale experimental study on the ambiguous utterance *Tell me about it* will be reported on.

KEYWORDS

Construction Grammar, usage-based, prosody, semiotic mode, forced-choice experiment

1 Introduction

Grammar deals with morphosyntactic patterns. True to this claim, the introductory sentence to the *Oxford Handbook of English Grammar* states that "'grammar' is used in the sense which encompasses morphology (the principles of word formation) and syntax (the system for combining words into phrases, clauses, and sentences)" (Aarts et al., 2019). Construction Grammar is no exception to this rule: Goldberg defines a grammatical construction as a "learned pairing of form with semantic or discourse function, including morphemes or words, idioms, partially lexically filled and fully general phrasal patterns" (Goldberg, 2006, p. 5). While Construction Grammar foregrounds the role meaning plays in forming grammatical structures, neither intonation nor prosody are explicitly mentioned. This is surprising to the extent that research at the prosody-meaning interface has a long tradition and intonation is acknowledged to fulfill grammatical functions (see e.g., Tench, 1996; Wells, 2006; Levis and Wichmann, 2015; Nolan, 2021). One of the reasons for separating prosody from grammar may have to do with the fact that even within prosody research, its grammatical function used to be downplayed, maintaining that "in practice it is usually context that disambiguates and the role of intonation is minimal" (Levis and Wichmann, 2015, p. 151), even though Wichmann and Blakemore (2006, p. 1,537) argue earlier that "[t]he choice of a rise or fall, or the placement of a pitch accent, may be as important a cue to speaker meaning as its phonetic realization." Rather, the so-called paralinguistic functions of prosody were foregrounded, i.e., its role in indicating emotions and attitudes (Féry, 2017, p. 7) and, indeed, the grammatical and the attitudinal functions of prosody are often interrelated (Gussenhoven, 2004).

Testifying to the "multimodal turn," the past decade has witnessed a rise in interest in multimodal Construction Grammar (see Section 2.2 below), i.e., an interest in constructions other than exclusively morphosyntactic ones. Part of the debate in this recent area of interest is the question of what defines a multimodal construction and, more specifically, which role prosody plays. While it seems uncontested that the combination of a morphosyntactic and a kinesic form might form a multimodal construction (see e.g., Ningelgen and Auer, 2017; Ziem, 2017; and other papers in Zima and Bergs, 2017; or in Uhrig, 2020), prosodic peculiarities of constructions are seldom addressed (notable exceptions include Lelandais and Ferré, 2019; Põldvere and Paradis, 2020). There is no *a priori* reason to exclude prosody from a constructional analysis, though; the only reason to do so seems to be the traditional misconception of prosody being something outside of the scope of grammar and, therefore, not worth any further consideration.

The aim of the present paper is twofold. First, it will show that prosody and morphosyntax can (and should) be considered independent semiotic modes (in the sense of Bateman et al., 2017), which independently can fulfill grammatical functions. Second, the paper will also show that the two semiotic modes can combine to form a multimodal construction (in the sense of Construction Grammar). The paper will proceed as follows: The main tenets of usage-based Construction Grammar and the notion of multimodal constructions will be introduced. Based on previous research, the paper will then argue that prosody and morphosyntax are independent semiotic modes by showing that they make use of different materiality and forms and that they independently contribute to the discourse semantics. It will then report on evidence that the two different modes may combine to form a multimodal construction using the results of a forced choice experiment.

2 (Usage-based) Construction Grammar and multimodality

In this section, the core assumptions of (usage-based) Construction Grammar and its relation to multimodality will be introduced. More specifically, the debate surrounding the notion of multimodal construction will be reviewed.

2.1 Constructions in Construction Grammar

Construction Grammar is no unified theory. For an overview of the different strands of Construction Grammar, Hoffmann and Trousdale (2013) is a useful resource. One of a few things all Construction Grammars have in common is that they consider the construction to be the core unit of language-related knowledge. A unit is considered a construction (C) "iff$_{def}$ C is a form-meaning pair $<F_i, S_i >$ such that some aspects of F_i or some aspect of S_i is not strictly predictable from C's component parts or from other previously established constructions" (Goldberg, 1995, p. 4). Figure 1 provides a schematic representation of a construction (taken from Croft and Cruse, 2004, p. 258). An example is the English idiom *Tell me about it*. Its component parts suggest (predict) that information is requested, but experienced language users know that it can also mean "'I'm well aware of that,' 'I agree;' 'you don't have to tell me'" (Tell, 2023). Since its meaning cannot be predicted from its component parts, it is a separate construction and must be learned. From such a perspective, idioms enjoy the same ontological status as words and more schematic constructions.

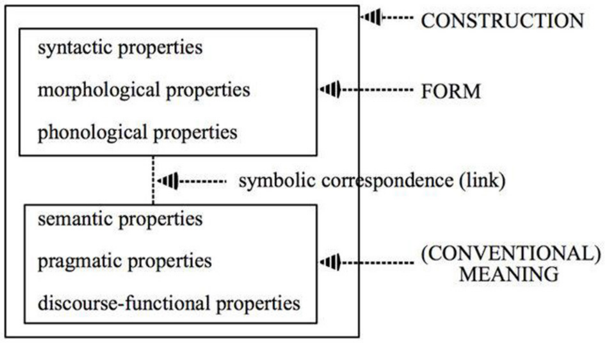


FIGURE 1
Schematic representation of a construction.

Usage-based approaches to Construction Grammar also consider predictable units to be constructions as long as they occur frequently enough so that they become entrenched in the language users constructicon, i.e., the mental repository of constructions (e.g., Bybee, 2006, 2013; Goldberg, 2006; Divjak, 2019). One example for this is the word *singer*. Even though its meaning "someone who sings" is perfectly predictable from its component parts, the verb *sing* and the derivational morpheme -*er*, the derivate *singer* is likely stored as a separate construction, because it is one of the 5,000 most frequent words in (written) English (Singer, 2023). Usage-based approaches to Construction Grammar further assume that the cognitive processes involved in language production and comprehension are domain-general and not specific to language. One of these domain-general cognitive processes is cross-modal association, which "allows humans to match up the phonetic (or manual) form experienced with properties of the context and meaning" (Bybee, 2013, p. 50), and which seems to be key in language learning (Imai and Kita, 2014; Dingemanse et al., 2015). An example of cross-modal association is sound symbolism, which is more pervasive in English than traditionally assumed. Sidhu et al. (2021) could show that sounds associated with roundedness (like /m/) more often than not denote round objects in English, while sounds associated with spikiness (like /k/) often denote spiky objects in English; an effect also known as the maluma/takete effect (Köhler, 1929).

2.2 Multimodal constructions

Constructions can be of any size, "including morphemes or words, idioms, partially lexically filled and fully general phrasal patterns" (Goldberg, 2006, p. 5) as well as argument and information structure constructions (see e.g., relevant chapters in Hoffmann and Trousdale, 2013; Hilpert, 2019; Hoffmann, 2022), but, evidently, the vast majority of constructions considered is of a morphosyntactic nature. This is surprising to the extent that usage-based Construction Grammar emphasizes language knowledge to emerge from the input language users get—and arguably this input commonly is multimodal. For instance, spoken language, i.e., the language infants are exposed to first, is inherently multimodal (Vigliocco et al., 2014; Feyaerts et al., 2017; Perniss, 2018), since speakers use gaze, gestures, facial expressions and other resources to convey meaning (see also Section 4.1 on the multimodality of *Tell me about it*). But also written language is often produced in multimodal situations (see e.g., Kress, 2000; van Leeuwen, 2014; Hiippala, 2017). Internet memes, for example, use written language and an image to convey their (conventionalized) meaning (Dancygier and Vandelanotte, 2017; Bülow et al., 2018). Despite these facts, multimodal constructional analyses are often noticeably absent from research in (usage-based) Construction Grammar.

In parallel to the multimodal turn in linguistics in general (see Stöckl, 2020), the past decade has also witnessed a growing interest in multimodal issues in Construction Grammar. One strand of research concerns itself with speech-embedded non-verbal depictions, i.e., gestures that may fill specific slots of constructions, such as Verb or Noun Phrase (see e.g., Clark, 2016; Ladewig, 2020; Hsu et al., 2021). Although not all of these studies position themselves in a Construction Grammar framework, their examples can be reanalyzed, like in Example (1):

(1) [MB was discussing a measure in a Mozart sonata] But then he writes "(*gazing at audience and singing*) dee-duh dum." That is very expressive.

(Clark, 2016, p. 325)

From a Construction Grammar perspective, the nonverbal depiction (i.e., *dee-duh dum*) fulfills the function of the object noun phrase in the transitive construction. Examples like these thus show that constructional slots need not be filled by morphosyntactic elements but can also be realized by other means.

Another strand of research discusses the possible existence of multimodal constructions. Ziem (2017) names four conditions under which a construction can be seen as multimodal, of which only the first two will be reviewed here, because they are central to the argumentation put forward in this paper.[1] The first condition states that

(a) A multimodal construction is a conventionalized pairing of a complex form that consists, at least, of a verbal element combined with a kinetic element (Ziem, 2017, p. 5).

In other words, a multimodal construction needs some kind of verbal form (with syntactic, morphological and/or phonological properties) and, necessarily, a kinetic element (like a manual gesture, a facial expression, or a particular gaze behavior) to be called such. Based on the representation of a construction (provided in Croft and Cruse, 2004, p. 258), Figure 2 depicts the representation of a multimodal construction.

A prime example for such a multimodal construction is the complex form of a deictic expression like *there* and a deictic gesture (like pointing, a head nod or directed gaze; Levinson, 2006), which, together, serve to identify a location in a given situation. This condition, however, may be and, as will be argued in this paper, is, in fact, incomplete. While a complex form might be a verbal plus a kinetic element, it might also be a verbal element plus a prosodic pattern. To show that the second combination is also a possible manifestation of a multimodal construction, it needs to be shown that morphosyntax and prosody are two different modes, each contributing independently to the meaning of the construction. Alternatively, it might be assumed that prosody is yet another aspect of unimodal constructions, on a par with their phonological properties. The review provided in Section 3 will rule out this alternative viewpoint.

The second condition Ziem (2017) puts forward runs as follows:

(b) Multimodal constructions manifest themselves either as inherently multimodal units or as entrenched cooccurrences of a verbal and a kinetic element (as opposed to constructions solely realized in a multimodal way).

This condition indicates that there are two kinds of multimodal constructions, which need to be kept distinct from incidental cooccurrences of e.g., a construction and a gesture (see also Hoffmann, 2017). The first kind of multimodal construction is inherently multimodal, i.e., it is non-predictable in some way. This holds for the combination of a deictic expression and a deictic gesture: The deictic expression remains incomplete in meaning (at least in some of the cases) unless it is used with deictic gesture. The second kind of multimodal construction follows from the usage-based premise that an expression can be fully predictable and still be a construction when it occurs with sufficient frequency. Schoonjans (2018), for example, could show that the German particle *einfach* cooccurs with a head shake in 24% in his corpus. Zima (2017) could show that [all the way from X PREP Y] is produced with a gesture in 80% of cases. And Uhrig (2022) could show that verbs of throwing are, on average, accompanied by a gesture in 54% of cases (with 66% for *fling* but only 42% for *lob*). Even though these corpus studies attest statistically significant cooccurrences of morphosyntactic and kinetic elements, they could only provide indirect evidence that this statistical significance can be equated with practical significance, i.e., show that these multimodal realizations constitute cognitive units. Therefore, in Section 4, the present paper will provide some evidence that language users actively make use of the prosodic mode to disambiguate (multimodal) constructions by reporting on a forced-choice experiment using the construction *Tell me about it*.

The present paper is not the first trying to bring together Construction Grammar and prosody. The past decade has also seen a rise in studies researching the prosody-syntax interface from a Construction Grammar perspective, but did so independently, i.e.,

[1] The other two conditions follow from the first two and therefore do not need explicit attention. The third condition specifies what should not be considered a multimodal construction (e.g., a construction only realized multimodally) and the fourth condition states that multimodal constructions need to be part of the constructional network of a language, i.e., a network that covers the relevant knowledge a speaker of that language needs for understanding.

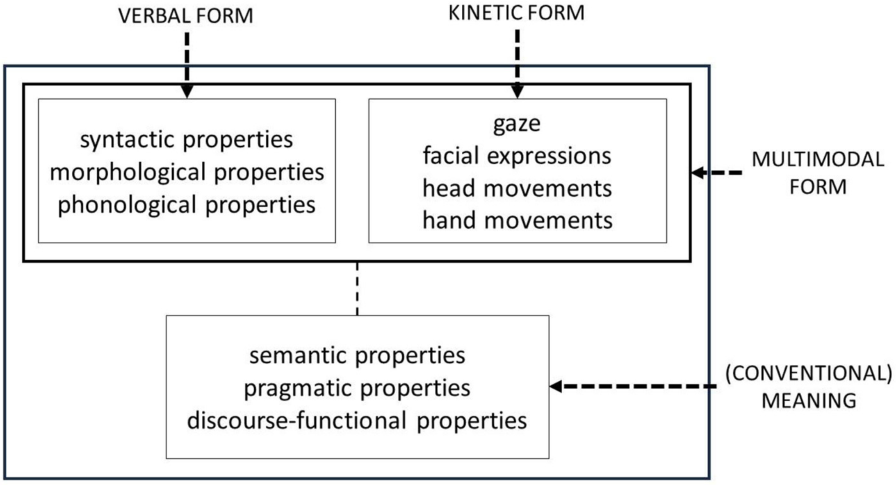


FIGURE 2
Schematic representation of a multimodal construction.

without referring to multimodal constructions. In the Introduction to their edited volume on Prosody and Construction Grammar, Imo and Lanwer (2020) summarize possible synergies. One possibility is the existence of prosodic constructions, i.e., assemblies of prosodic features that convey a particular meaning (relatively) independent of the words that are used with it. These prosodic constructions combine with morphosyntactic constructions in an *ad hoc* manner if their functions are compatible. Prosodic constructions have been proposed for French (Marandin, 2006), Persian (Sadat-Tehrani, 2010), Spanish (Gras and Elvira-García, 2021), and English (Ward, 2019). Another possibility is that prosodic properties, if recurring, can be part of the formal side of the (unimodal) construction. This was proposed for the reactive *what-x* construction (*What mince pies?*), which reacts to something in the preceding turn by another speaker and needs to be prosodically integrated (Põldvere and Paradis, 2020). And, finally, a third possibility is that prosody and morphosyntax interact in a meaningful way such that a construction would be incomplete without considering both components and none of the two components constitute independent constructions. This seems to be the case for German appositive structures (e.g., *der Spitzenkoch Tim Mälzer*, English *the top chef Tim Mälzer*), as evidenced in Lanwer (2020). Even though this is not made explicit, this possible relation between prosody and Construction Grammar fits the definition of a multimodal construction with the only exception that "kinetic" form needs to be replaced by "prosodic" form. Figures 3–5 summarize all possible configurations.

In a nutshell, the present paper aims to show that there are multimodal constructions that consist of a syntactic and a prosodic form, which combine to convey one meaning. To do so, evidence for a prosodic mode (in English) will be reviewed to show that, in principle, prosody and morphosyntax (or rather the phonological properties of morphosyntactic elements) are two different modes. Moreover, a forced-choice experiment will be reported on, which shows that certain prosodic forms are not just used incidentally, but that they are part of language users' knowledge.

3 Evidence for a prosodic mode in English

There are many definitions of the term mode and some of them equate mode with sensory channel. Such a, often pre-theoretical, notion of mode might be one of the reasons why prosody has been largely neglected in usage-based, multimodal approaches to Construction Grammar. From such a view, prosody and spoken language belong to the same mode and, thus, need not be part of multimodal analyses. The present paper, however, will use the notion of semiotic mode, which is prevalent in multimodality research. More specifically, the paper will make use of the definition of semiotic mode as proposed by Bateman and colleagues (Bateman, 2011, 2022; Bateman and Wildfeuer, 2014; Bateman et al., 2017).

Bateman defines a semiotic mode as "a three-way layered configuration of semiotic distinctions developed by a community of users in order to achieve some range of communicative or expressive tasks" (Bateman, 2022, p. 68). The first layer of the semiotic mode is the material substrate, i.e., "the 'stuff' which is used when making meaning" (Bateman and Wildfeuer, 2014, p. 181). In other words, semiotic agents manipulate the material to communicate. The second layer is the form side of the mode. The form consists of categories derived from the (noisy) material that are, by convention, used to distinguish meanings. These forms can be simple or complex. And, finally, the third layer of the semiotic mode is that of discourse semantics, i.e., the meaning contribution of the mode in relation to its surroundings. The following subsections will show if and to what extent (spoken) morphosyntax and prosody differ along these lines.

3.1 The material substrate

From an articulatory perspective, the material substrate of spoken English morphosyntax is part of introductory knowledge

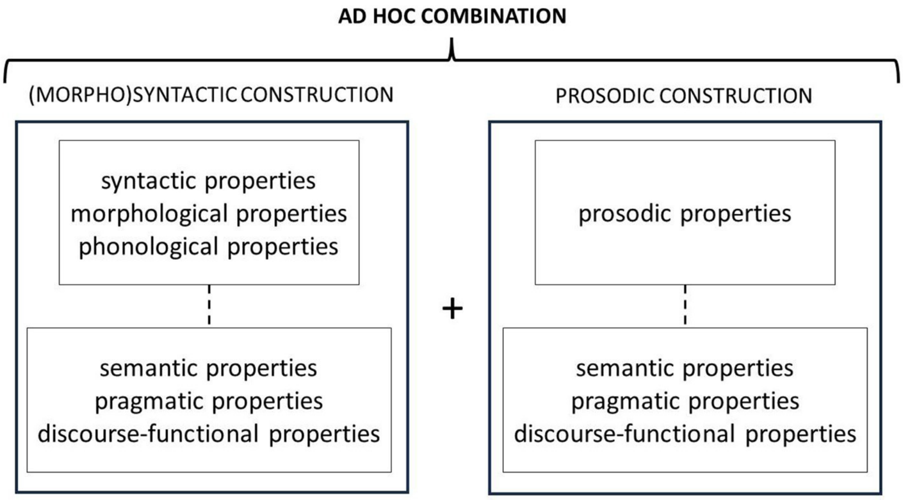


FIGURE 3
Schematic representation of an *ad hoc* relationship between prosodic and (morpho) syntactic constructions.

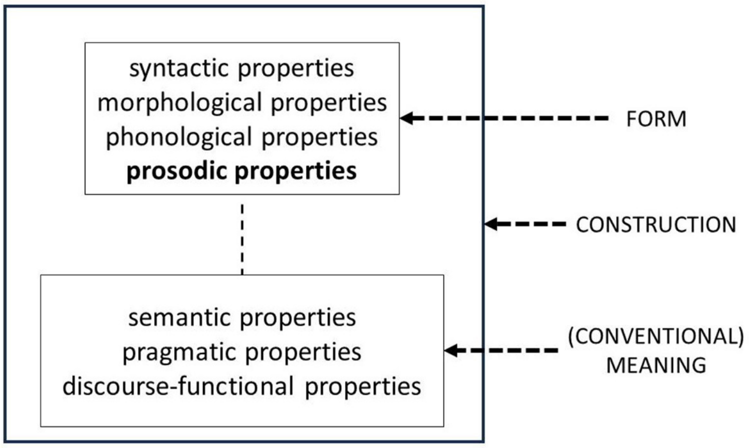


FIGURE 4
Schematic representation of a unimodal construction with prosodic properties.

in linguistics. Speakers use the air stream coming from the lungs and manipulate this air stream with the help of different, active and passive, articulators to create sounds. One main active articulator is the vocal folds, which can produce voiced sounds when vibrating and voiceless sounds when not vibrating. The other articulators of English sounds are mainly found in the oral cavity: the lips, the teeth, the tongue, the alveolar ridge, the hard and the soft palate (also called velum) as well as the uvula (depending on the variety of English spoken). Acoustically, this manipulation of the airstream results in different shapes of the sound waves produced. For example, plosive sounds are characterized by a silent period and a sudden release burst, fricatives by a strong turbulence noise and vowels by energy peaks at certain frequencies (also known as first and second formants), to name but a few.

The articulatory mechanisms behind prosodic features in English (to be discussed below) partially overlap with that of the sounds of English. The most central prosodic features—pitch, loudness, and duration—are manipulated largely with the help of the diaphragm and the vocal folds. The diaphragm is a large muscle below the lungs that controls breathing and thus, the airstream. The greater the airflow, the louder the speech tends to get. The diaphragm is also involved in producing (English) speech sounds, because, when there is no airflow, no sounds can be produced.[2] Technically, speakers may also "speak from their throats," i.e., without support from the diaphragm, but even that has respiratory constraints. Still, even though the diaphragm is involved in the production of speech sounds, it does not have an influence on the perception of these sounds as phonemes. A/l/is a/l/, no matter whether it is loud or quiet. Acoustically, with greater airflow, the pressure the sound signal exerts on the surrounding particles is

[2] Languages other than English have non-pulmonic sounds, i.e., sounds where the airflow does not come from the lungs, but these will not be considered here.

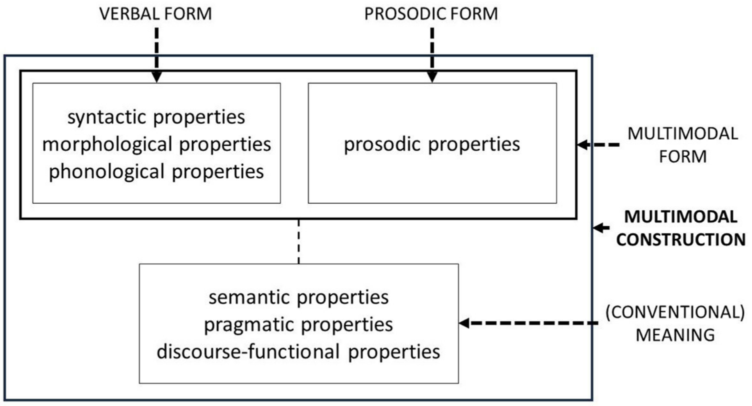


FIGURE 5
Schematic representation of a multimodal construction made up of a morphosyntactic and a prosodic form.

higher. The other main articulator in prosody is the vocal folds, which are responsible for pitch production. The speed with which they vibrate correlates with the fundamental frequency (f_0) of the sound produced. The faster they vibrate, the higher the sound is perceived. As outlined above, the vocal folds are also involved in sound production. However, even though the articulator is the same, it does two different things here. For sound production, it is important to either let the vocal folds vibrate or not. For pitch, what matters is the speed with which they vibrate. From an acoustic perspective, higher frequency of vibration causes the sound waves to oscillate faster, too.

All in all, what can be seen from this necessarily brief overview is that sounds (as the building blocks of spoken morphosyntax) and prosodic features are produced by different parts of the articulatory system. This means that they can be (and are) manipulated independently in the meaning-making process and, thus, also can take on different forms.

3.2 Form

Regarding the form of spoken English morphosyntax, the paper will only consider phonological categories, since these are most central for the present argument. The phonological features that serve meaning-distinguishing purposes in English are the state of the glottis, the manner of articulation, and the place of articulation for consonants, and the positioning of the tongue and duration for vowels. For English vowels, further meaning-distinguishing features have been proposed, either in addition or substituting duration, namely muscular tension and position of the lips. In any case, features like these enable language users to distinguish categories such as /b/ and /p/ (state of the glottis), /b/ and /m/ (manner of articulation), /b/ and /d/ (place of articulation), /i:/ and /u:/ (position of the tongue) as well as /i:/ and /ɪ/ (duration, but also position of the tongue).

For prosody, features that serve meaning-distinguishing purposes include, at least, the “big three” pitch (the perceptual correlate to fundamental frequency), loudness (the perceptual correlate of the pressure of the sound signal), and aspects of timing (such as speaking rate, articulation rate or pauses). These features enable the language user to perceive categories such as rising and falling intonation (pitch), loud and quiet speech (loudness) as well as fast and slow speaking tempo (timing). These three often work together to form prosodic constructions, i.e., configurations of prosodic forms that convey a particular meaning independent of the words used (see Section 3.3. below for examples). There are further prosodic features, such as voice quality (nasality, creakiness) and articulatory precision, but these seldom serve meaning-distinguishing functions on their own. In sum, there is some overlap regarding the meaning-distinguishing features of spoken morphosyntax and prosody, since (vowel) duration and timing are both time-related features, but other than that, the features can clearly be distinguished from one another. What is more, even though vowel duration and timing seem to correlate, language users are able to distinguish the two nonetheless. Just consider a word like *bit*. Its vowel, /ɪ/, is short in duration, but the meaning of the word does not change if it is pronounced in a slow manner (which is the case, of course, because no two words in English are ever distinguished by vowel duration alone) as long as the contrast with other vowels of a similar quality is maintained.

An interesting exception might be stress placement. There are words in English that only differ by word stress, e.g., *differ* /’dɪfə/ or /’dɪfər/ and *defer* /dɪ’fə:/ or /dɪ’fər/. The acoustic correlates of stress in English include, among others, pitch, loudness and timing (see e.g., Fry, 1955, 1958; Lieberman, 1960), i.e., the “big three” mentioned above. Examples like *differ* and *defer* blur the lines between meaning-distinguishing features that are relevant for morphosyntax and those for prosody. Therefore, one could treat them as counterevidence that prosody is an independent mode because a prosodic configuration that language users perceive as word stress serves morphosyntactically relevant functions. Likewise, it could be argued that words like *differ* and *defer* are, in fact, multimodal constructions combining a phonological (e.g., /dɪfə/) and a prosodic form (e.g., /’σσ/) for *differ*. It is outside the scope of the present paper to provide evidence for one or the other claim. Still, the argument put forward in the following clearly favors the second option.

3.3 Discourse meaning

From a Construction Grammar perspective, all morphosyntactic units of interest, i.e., constructions, carry meaning per definition (although this is not uncontroversial, see e.g., Fillmore et al., 2012 on constructions without meaning). Therefore, there is no need to discuss the meaning of these.

The more interesting question is rather whether prosodic forms, independent of the words that are used with them, carry meaning. There is, in fact, quite some evidence for the existence of prosodic constructions. Prosodic constructions have been identified for Spanish (Elvira-García, 2019; Gras and Elvira-García, 2021), German (Neitsch and Niebuhr, 2019; Niebuhr, 2019), French (Marandin, 2006), Persian (Sadat-Tehrani, 2010), and most notably for the present purposes, English (Ward, 2019). One of the prosodic constructions attested for English, the *consider this* construction, will be reviewed in more detail, because it is one of the constructions that is understood best. This prosodic construction was first described in Liberman and Sag (1974) and is attested both experimentally (Kurumada et al., 2012) and with the help of corpora (Hedberg et al., 2003; Ward, 2019). Its formal features are illustrated in Figure 6. While most of its formal descriptions focused on the pitch movements only, recent advances show that it consists formally of three parts: The first is a region that is high-pitched, loud and slow, to be seen on the word *LOOKS* in Figure 6. The second is a region of level pitch, which can be seen on *like a ze-* in Figure 6. And, third, another high-pitched region, visible on the last syllable *-bra* in Figure 6 (Ward, 2019, p. 5–24). Functionally, it marks some kind of contradiction or contrast, a piece of information that is offered to the hearer for further consideration. Thus, the syntactic string *It looks like a zebra* uttered with the prosodic pattern described above implies that even though the animal in question might resemble a zebra, it is actually some other animal (Kurumada et al., 2012). There is compelling evidence that this form-function pairing is indeed conventionalized in American English: Corpus studies suggest that this prosodic form is more often than not used with contradictions (Hedberg et al., 2003; Ward, 2019) and experimental evidence suggests that language users favor a "no zebra" interpretation when presented with an utterance like depicted in Figure 6 (Kurumada et al., 2012). What is more, Liberman and Sag (1974) even argue that "without having any idea of the content of his utterance, we know from the melody performed ... that [the speaker] objects in some way" (422), i.e., that the prosodic form has an independent meaning. This independent contribution to the discourse semantics of prosody is probably the most convincing piece of evidence that prosody is an independent semiotic mode.

4 Entrenching prosodic information: *Tell me about it*

Section 3 argued that prosody is best seen as an independent semiotic mode. For the discussion on the relation between prosody and morphosyntactic constructions this means that prosodic properties cannot be analyzed on a par with other properties of morphosyntactic construction but need independent consideration. Section 3.3, in particular, has shown that there are prosodic constructions, like the *consider this* construction, that may combine with morphosyntactic constructions in an *ad hoc* manner to form a multimodal construct. In what follows, the paper will present some evidence for a genuinely multimodal construction, i.e., a construction with both entrenched prosodic and morphosyntactic properties. The construction under consideration is called stance-related *Tell me about it* and will be contrasted with another, formally similar construction, i.e., requesting *Tell me about it*.

4.1 Requesting and stance-related *Tell me about it*

Formally, requesting and stance-related *Tell me about it* (henceforth TMAI) are morphosyntactically similar. While formal variations for the stance-related construction can be found (e.g., *Tell me more* or *Tell me more about it*), these are rare and *Tell me about it* seems to be the preferred variant as this is the only form that is listed in dictionaries (e.g., in the Oxford English Dictionary Online, Tell, 2023). Functionally, the two TMAI constructions fulfill different, non-overlapping functions. Requesting TMAI is used to request information as is illustrated in Example (2).[3]

(2) "sci-fi thriller" (simplified)
A: I know she also has a sci-fi thriller. Arrival.
B: Uh-huh.
A: Tell me about it. Is it worth seeing?
B: Absolutely.
(2016-09-25_0832_US_KNBC_Access_Hollywood, 29:41-29:48).

In Example (2), speaker A introduces a referent, i.e., a science fiction thriller called *Arrival*. After speaker B's brief backchannel, speaker A encourages speaker B to provide more information on this film using TMAI and specifies the preferred continuation to be an evaluation (i.e., *Is it worth seeing?*). Speaker B then provides the requested information. As can be seen in this example, requesting TMAI usually initiates speaker transition. This transition need not occur directly after issuing TMAI, but constitutes what Sacks et al. (1974) call a transition-relevance place. Moreover, the next turn is expected to be an informing sequence, providing some more information on the referent that was introduced shortly before.

Stance-related TMAI fulfills completely different functions as is illustrated in Example (3).

(3) "we're all getting older" (simplified)
A: We're getting older. We're all getting older. So...
B: ((laughs)) T- Tell me about it.
A: ((laughs))
(2021-11-26_0600_US_KNBC_Dateline_NBC, 03:39-03:44).

[3] All examples of TMAI come from the *NewsScape Library of International Television News*, an archive of televised discourse (Steen and Turner, 2013). At the end of each example, the name of the source file and the relevant times are provided. Video snippets of the examples are provided on OSF: https://osf.io/2sq7h/?view_only=746f3703bbde4236b832b34234d51beb.

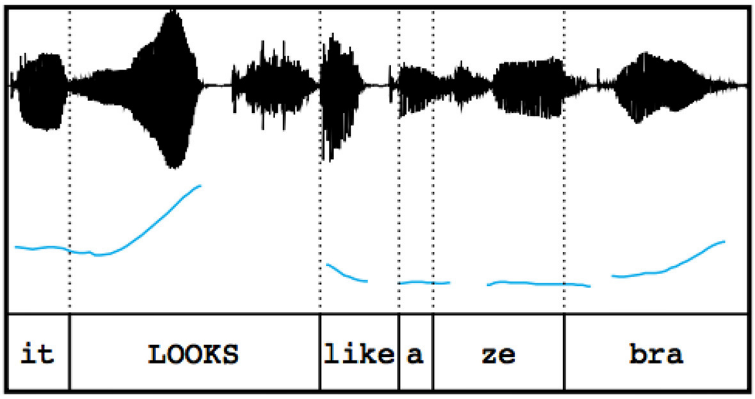


FIGURE 6
Waveform and pitch contour of the "consider this" construction (taken from Kurumada et al., 2012).

In Example (3), speaker A makes an observation (*we're getting older*), which many people find saddening. This seems also to hold for speaker A since he repeats this utterance, slightly modifying it (*we're all getting older*). Speaker B reacts to this observation, at first, with laughter and then with stance-related TMAI. This construction expresses an affective stance, i.e., a saddening view on aging. Likewise, it expresses epistemic authority. Speaker B is, apparently, older than speaker A and thus claims to be more knowledgeable person on this matter. Crucially, stance-related TMAI neither necessitates speaker transition nor an informing sequence. Speaker A reacts with laughter to speaker B uttering TMAI and the conversation is cut at this point.

It could be argued that both TMAI constructions are ambiguous and are only disambiguated in predictive context. However, in a corpus study using the multimodal *NewsScape Library of International Television News* (Steen and Turner, 2013), Lehmann (2023) showed that stance-related TMAI, when compared to requesting TMAI, is produced, more often than not, with raised eyebrows, averted gaze, smiling, some kind of head movement (often nods, shakes or tilts) and a slower speaking rate. This is illustrated with frame grabs of Example (3), which are provided in Table 1.

As can be seen Table 1, before uttering stance-related TMAI, speaker B looks at his interlocutor, already smiling. At the onset of TMAI, he turns his head (line 2) to the left and avoids eye contact with the recipient. In addition, he raises his eyebrows and continues smiling (see also line 3). Only after finishing uttering TMAI, on the last syllable, he turns his head orientation and his gaze back toward his interview partner. The duration of TMAI in Example (3) is 667 ms, which corresponds to a speaking rate of 7.4 syllables per second. This is very close to the mean speaking rate of stance-related TMAI in face-to-face interactions, which is 7.48 syllables per second, whereas requesting TMAI is faster in these contexts, with a speaking rate of 8.44 syllables per second (see Lehmann, in press).

All of these visual as well as prosodic properties of stance-related TMAI were shown to be statistically significant (Lehmann, 2023), but as was argued above, some Construction Grammarians claim that statistical significance need not be equated with practical significance. Therefore, both visual and prosodic properties of TMAI were put to the test in a forced choice experiment to provide evidence that language users indeed draw on these properties when interpreting an instance of TMAI.

4.2 Putting the multimodal properties of *Tell me about it* to the test

4.2.1 Method

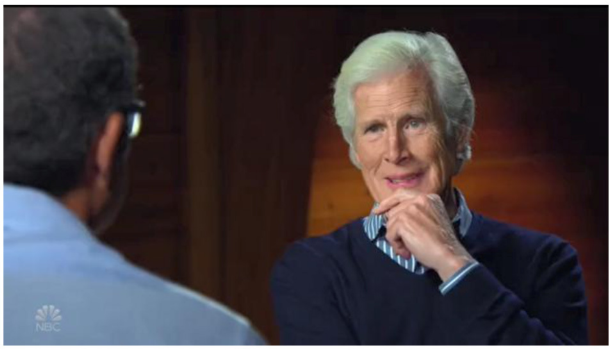
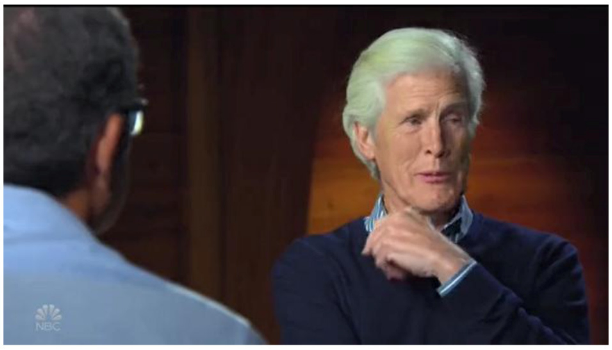
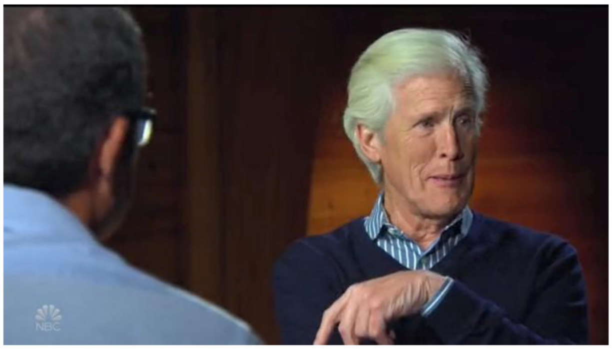
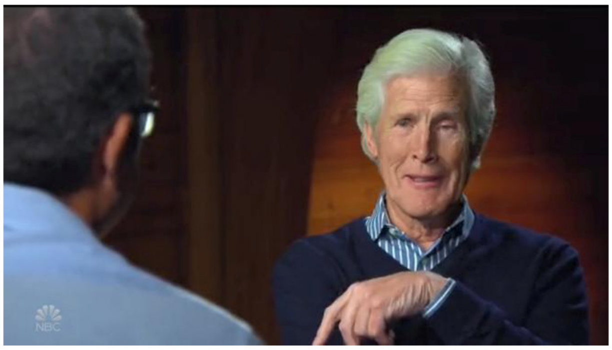
4.2.1.1 Participants

The participants in this experiment were 25 adult native speakers of American English, who were recruited via Prolific Academic (Palan and Schitter, 2018). They were rewarded £4.50 for their participation. In addition, 18 adult advanced learners of English participated. These were students of the study program *English-speaking Cultures* at the University of Bremen, Germany. To be admitted to this study program, students need to have a command of English at level B2 ("independent user") of the Common European Framework of Reference for Languages (Council for Cultural Co-operation, Education Committee, and Modern Languages Division, 2001), but many of them self-reported to know English on a C1 level ("proficient user"). They participated for course credit.

4.2.1.2 Procedure

The participants were requested to complete an online forced choice experiment, which had been designed with SoSci Survey (Leiner, 2021). In the instructions to this experiment, the participants were introduced to the two uses of TMAI, named *requesting information* and *ironic rejoinder*. This was done to make sure that the non-native speaker understand the task (in case they did not know TMAI could also be used in a stance-related way) and to introduce the two response options in the experiment. The label *ironic rejoinder* was preferred over the label *stance-related* in the experiment because the *Oxford English Dictionary* defines stance-related TMAI this way (Tell, 2023). The participants were told that they would see and/or hear a speaker uttering TMAI and that their task was to guess whether this utterance is requesting information or an ironic rejoinder.

TABLE 1 Frame grabs of an extract of example (3).

Line	Speaker	Utterance	Frame grab
1	A	So...	
2	B	t-	
3		Tell me about	
4		It	

4.2.1.3 Stimuli

The experiment consisted of 69 stimuli in total. All of these were selected observations of the corpus study from Lehmann (2023). These observations were presented in four different conditions. In the first condition, called "context condition," the participants were presented with TMAI with what was considered sufficient sequential context to disambiguate TMAI with the help of this context. This served as the reference condition. In the second

TABLE 2 Overview on the stimuli used in the experiment.

Condition	Description	Anticipated interpretation
Context	TMAI embedded in sequential context	Requesting ($N = 5$)
		Stance-expressing ($N = 5$)
Multimodal	Stand-alone TMAI Visual and acoustic information provided	Requesting ($N = 5$)
		Stance-expressing ($N = 4$)
		Ambiguous ($N = 9$)
Visual	Stand-alone TMAI No acoustic information Pace slowed down	Requesting ($N = 5$)
		Stance-expressing ($N = 5$)
		Ambiguous ($N = 11$)
Acoustic	Stand-alone TMAI No visual information	Requesting ($N = 5$)
		Stance-expressing ($N = 4$)
		ambiguous ($N = 11$)

condition, called "multimodal condition," the participants could both hear and see a speaker uttering TMAI, but without further sequential context. In the third condition, called "visual condition," the participants saw a speaker uttering TMAI, but they could not hear this person. Since these video snippets were extremely short with less than a second and some online video players have a time lag, the videos were played in slow motion. The participants were informed about this. Furthermore, to facilitate speaker identification in case there was more than one speaker visible, the videos were edited to such an extent that only the speaker of TMAI was visible. Finally, in the fourth condition, called "acoustic condition," the participants were provided with an audio recording of a speaker uttering TMAI only. Within, but not between these conditions, stimuli rotated.

The stimuli were further selected regarding their anticipated interpretation. The statistical model that was fitted for the corpus data in Lehmann (2023) makes clear predictions about how participants should interpret these stimuli, if the results were of practical significance. Thus, stimuli were selected according to the visual and/or prosodic features that the speakers used during the utterance. That is, some stimuli were selected as either prototypically requesting or stance-related uses of TMAI, when they displayed the properties that the statistical model predicted. Vice versa, some of the stimuli were selected as ambiguous stimuli when they displayed conflicting properties, e.g., when the speaker raised their eyebrows (a property of stance-related TMAI) but continued looking at the recipient (a property of requesting TMAI). Table 2 gives an overview on the stimuli used in the experiment.

4.2.2 Statistical analysis

The results of the forced choice experiment were analyzed with R (R Core Team, 2022). With the help of the *glmer* function of the lme4 package (Bates et al., 2015), a generalized linear mixed-effects model was fitted. The correctness of the response (i.e., whether the response was in line with the actual construction) was treated as the dependent variable. Initially, participant, language proficiency, stimulus, and construction were entered as random intercepts, while condition and anticipated interpretation were entered as fixed effects. This led to problems with convergence due to its complexity. An inspection of the initial model with the *summ* function of the jtools package (Long, 2022) showed that language proficiency and participant were negligible effects and were, thus, removed from the model. No problems with convergence occurred thereafter. The *summ* function was used to summarize the fitted model, including the computation of confidence intervals, and the ggplot2 package (Wickham et al., 2023) as well as the sjPlot package (Lüdecke, 2023) were used to visualize the fitted model.

4.2.3 Results

Figure 7 shows the overall distribution and central tendencies of correct responses for the different stimuli across conditions.

Figure 7 suggests that, overall, the participants were successful at guessing the meaning of TMAI based on visual and/or acoustic cues alone, given that the median ratio of correct guesses for the unambiguous stimuli is higher than 0.75. Figure 7 also suggests that, when compared to the context condition, participants seemed to have difficulties with the ambiguous stimuli, but neither the requesting nor the stance-related ones, except for five stimuli which score lower than 0.75, three of which in the visual condition and two in the acoustic condition.[4] In general, participants perform worse in the visual and the acoustic condition than in the multimodal condition. In these two conditions, the ambiguous stimuli seem to pose the greatest difficulties to the participants, as expected.

Table 3 provides a summary of the fitted model and Figure 8 shows the odds ratios of the model terms (condition and anticipated interpretation).

With a pseudo-R^2 of 0.64 for the total effects and a pseudo-R^2 of 0.36 for the fixed effects, the model summarized in Table 3 explains a good amount of variance in the responses obtained. It shows that the participants were significantly worse at guessing the meaning of TMAI in the multimodal (with $p = 0.04$, OR $= 0.13$), visual (with $p < 0.001$, OR $= 0.02$) and acoustic condition (with $p < 0.001$, OR $= 0.002$) when compared to the context condition. It further shows that there is no significant difference between guessing requesting and stance-related TMAI correctly (with $p = 0.33$, OR $= 2.09$), but the ambiguous stimuli contribute to the model with borderline significance (with $p = 0.06$, OR $= 0.35$), suggesting that most incorrect guesses were due to the ambiguous stimuli, but not entirely.

[4] There seem to be at least two reasons why the participants scored low in correctness for these prototypical stimuli. One reason might be the timing of TMAI and the visuals. That is, for some visual stimuli, some important visual displays (gaze aversion, raised eyebrows, and smiling) occurred right before, but not during the speaker uttered TMAI. This non-synchrony might have affected the speakers' choices. Another reason might be that the model reported in Lehmann (2023) is incomplete. It seems that, while the duration of TMAI is a good predictor, it is not the only one.

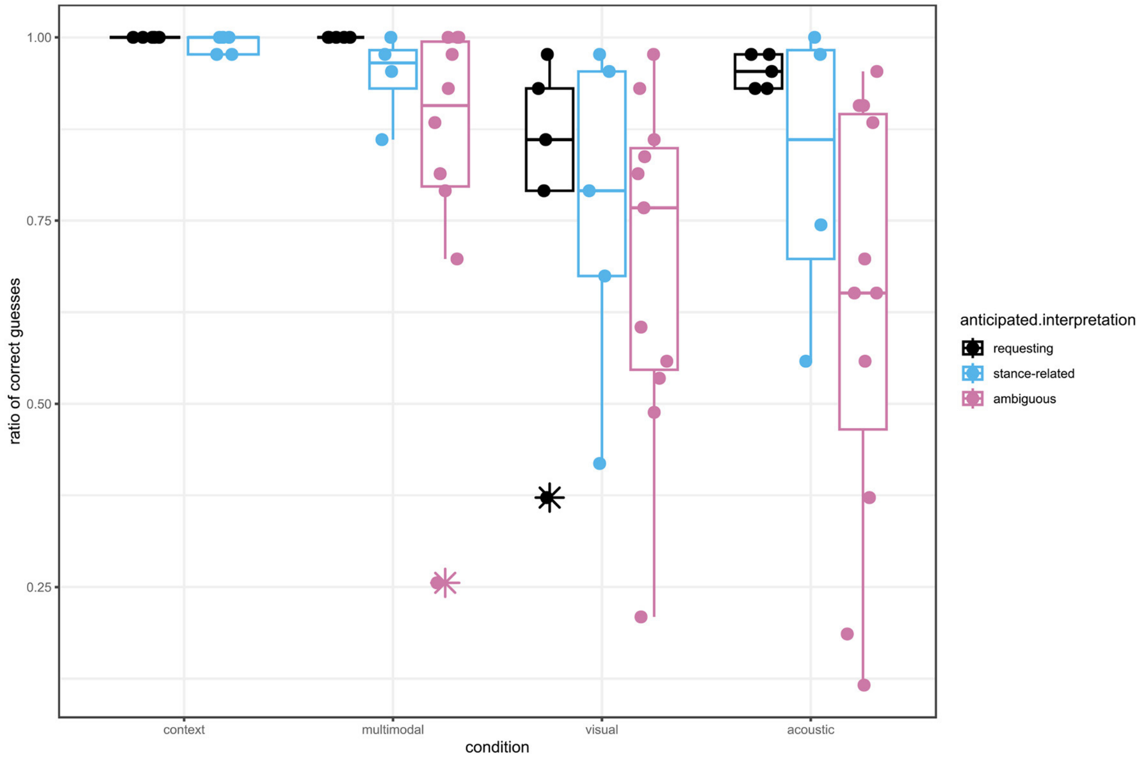


FIGURE 7
Grouped boxplots with jitter of correct responses regarding the anticipated interpretation across conditions. The asterisk indicates outliers.

4.2.4 Discussion of the forced-choice experiment

The experiment reported above shows that prosody alone can disambiguate TMAI if the prosodic features that are associated with the construction are displayed, i.e., the speaking rate in this case. If TMAI is ambiguous regarding its speaking rate and hearers lack other pieces of information, they seem to have difficulties in guessing its meaning. Vice versa, if the speaker produces TMAI with a slower speaking rate, hearers are more likely to understand this as stance-related TMAI, even if there are no further features available. Interestingly, the results also suggest that hearers use prosodic information alone to disambiguate TMAI about as accurate as they use visual information alone. This observation might suggest that the strength of association between prosodic properties and the construction is comparable to the one between visual properties and the construction.

Technically, these observations can be explained in two ways. One explanation is that slow speaking rate is an independent prosodic construction. Niebuhr (2010), for example, has shown that lengthened consonants correlate with negative sentiment in German. The same could be true for English stance-related TMAI. Informal observations of TMAI, however, suggest that it is not the lengthening of the consonants alone that result in a slower speaking rate, but also the lengthening of the vowels. At the same time, speaking rate alone does not explain all the findings observed in the experiment. There are quite a few stimuli that were neither slow nor fast (i.e., ambiguous), which posed no difficulties to the participants. This suggests that there might be more, albeit undetected, prosodic features associated with TMAI. Given that, it is possible that there is a (complex) prosodic construction that is often used with stance-related TMAI, but, at the moment, there is only scarce evidence for that. The other way to explain the findings of the experiment is to assume that the slow speaking rate is part of the stance-related construction, forming a multimodal construction. If there is, indeed, no prosodic construction that can be identified, and given that prosody is a mode, then stance-related TMAI must be considered a multimodal construction with morphosyntactic and prosodic (and, possibly, visual) features. Even if future studies show that there is a prosodic construction such as "slow speaking rate," both the frequency with which it is used with stance-related TMAI and the apparent use of this construction to disambiguate TMAI would speak in favor of treating TMAI as a multimodal construction from a usage-based perspective.

5 Conclusions

The present paper had two objectives. The first objective was to show that prosody and morphosyntax are two independent semiotic modes with distinguishable differences in material and form as well as independent contributions to the discourse semantics. It could be shown that the aspects of the sound stream that are relevant for spoken morphosyntax are not the same as the aspects that are relevant for prosody. Using these different aspects, hearers transform the input from the sound stream to either arrive at categories like /p/, /m/ or /e/ (spoken language) or high pitch, loud speech, and/or fast

TABLE 3 Summary of the fitted model for correct responses.

Model info:					
Observations: 3010					
Dependent Variable: correctness					
Type: Mixed effects generalized linear regression					
Error Distribution: binomial					
Link function: logit					
Model fit:					
AIC = 1,997.97, BIC = 2,046.05					
Pseudo-R² (fixed effects) = 0.36					
Pseudo-R² (total) = 0.64					
Fixed effects:					
	Est.	2.5%	97.5%	*z* val.	*P*
(Intercept)	5.87	3.46	8.27	4.79	<0.001
Multimodal	−2.03	−3.97	−0.10	−2.06	0.04
Visual	−4.09	−5.97	−2.20	−4.25	<0.001
Acoustic	−3.79	−5.69	−1.89	−3.91	<0.001
Stance-related	0.74	−0.74	2.22	0.98	0.33
Ambiguous	−1.06	−2.16	0.04	−1.89	0.06
Random effects:					
Group	Parameter	Std. dev.			
Stimulus	(Intercept)	1.25			
Construction	(Intercept)	0.97			
Grouping variables:					
Group	# groups	ICC			
Stimulus	70	0.27			
Construction	2	0.16			

speech (prosody). These categories are then combined to form meaningful structures like *It looks like a zebra* (spoken language) or (contextually meaningful) assemblies conveying "consider this" (prosody), and they do so largely independent of one another. Since spoken language and prosody differ in all three layers of the semiotic mode, they must be considered independent. For constructional analyses, this means that prosody cannot be represented on a par with other, morphosyntactic and phonological, properties. Rather, it needs its own place. This place could take on the form of a prosodic construction (in case the prosodic configuration has an independent meaning) or of being part of a multimodal construction (in case the prosodic configuration has no independent meaning). Such a view on prosody strengthens the multidimensional network approach to language-related knowledge, which assumes that constructions are interrelated by various kinds of associations (Diessel, 2023). Prosodic constructions as well as multimodal constructions are prime examples of such a network of (cross-modal and multimodal) associations.

The second objective of the present paper was to provide evidence for a multimodal construction consisting of, at least, a morphosyntactic and a prosodic form. Both corpus and experimental evidence suggest that the stance-related use of *Tell me about it* is a likely candidate for such a multimodal construction. Regarding its prosodic form, stance-related *Tell me about it* is slower in tempo than its requesting counterpart. When language users are provided with nothing but this difference in tempo (i.e., they lack other clues like sequential context or visuals), they use this prosodic feature to disambiguate *Tell me about it*. In other words, this knowledge on the two uses of *Tell me about it* must be stored in the language users' minds in some way. Stance-related *Tell me about it* thus fulfills Ziem's second condition of multimodal constructions, because it cannot be considered a construction that is "solely realized in a multimodal way," but the paper has shown that it is an entrenched cooccurrence of a verbal and a prosodic form. In conclusion, the evidence presented in this study on *Tell me about it* is strongly suggestive of the existence of multimodal constructions. As a consequence, the role

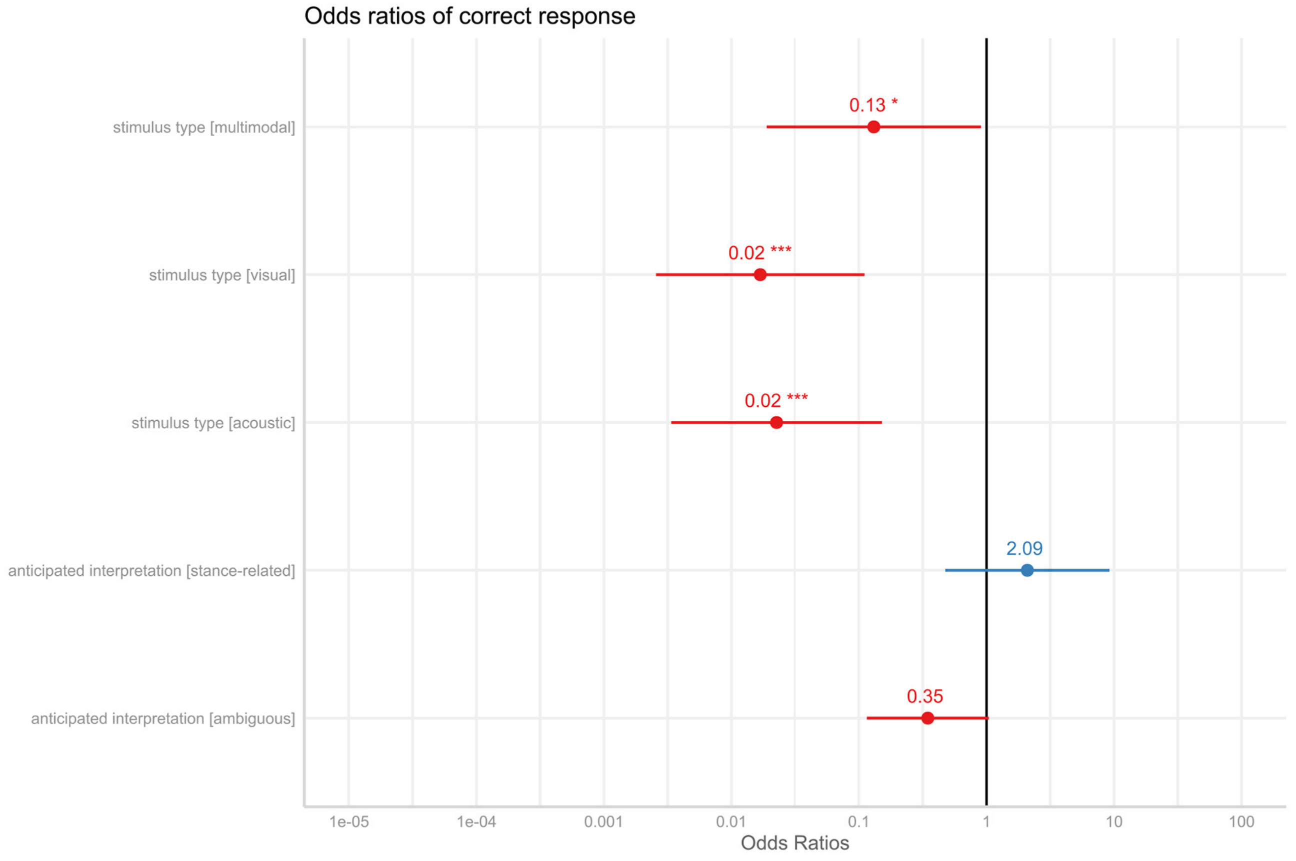


FIGURE 8
Odds ratios of correctness of response [$p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***)].

prosody plays in forming them needs more systematic attention in constructional analyses.

From a methodological perspective, the present paper could show that a triangulation of corpus and experimental evidence is valuable because it was able to shed light on both the production and the comprehension side of language and, in doing so, draw a complementary picture of prosody and multimodal constructions. However, the present study suffers from obvious limitations that require further systematic attention in future studies. One limitation is the low number of participants in the forced-choice experiment and the missing demographic information. From a usage-based perspective, the constructional network (including multimodal and prosodic constructions) is dynamic and, therefore, can vary for certain demographic groups. This aspect is not reflected in the present study and needs to be addressed in the future. In addition, future research also needs to address the role prosody plays in the constructional network in more detail. Studies that explore prosodic and multimodal constructions could identify the exact (inter)relations and associations between different types of constructions and, thereby, provide an answer to the question if multimodality is a central or a peripheral aspect of grammar.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://osf.io/2sq7h/?view_only=746f3703bbde4236b832b34234d51beb.

Ethics statement

Ethical approval was not required for the study involving human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants in accordance with the national legislation and the institutional requirements.

Author contributions

CL: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer 491466077.

Acknowledgments

My thanks go to John Bateman. John was the best mentor you can imagine during my time at the University of Bremen. Without him, this paper would be less rigorous and less advanced in almost every respect.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aarts, B., Bowie, J., and Popova, G. (2019). "Introduction," in *The Oxford Handbook of English Grammar*, eds. B. Aarts, J. Bowie, and G. Popova (Oxford: Oxford University Press), 1.

Bateman, J. (2011). "The decomposability of semiotic modes," in *Multimodal Studies: Exploring Issues and Domains*, eds. K. O'Halloran and B. Smith (New York, NY: Routledge), 17–38.

Bateman, J. (2022). Growing theory for practice: empirical multimodality beyond the case study. *Multimodal Commun.* 11, 63–74. doi: 10.1515/mc-2021-0006

Bateman, J., and Wildfeuer, J. (2014). A multimodal discourse theory of visual narrative. *J. Pragmat.* 74, 180–208. doi: 10.1016/j.pragma.2014.10.001

Bateman, J., Wildfeuer, J., and Hiippala, T. (2017). *Multimodality: Foundations, Research and Analysis – A Problem-Oriented Introduction*. Berlin: De Gruyter Mouton.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using Lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bülow, L., Merten, M. L., and Johann, M. (2018). Internet-Memes Als Zugang Zu Multimodalen Konstruktionen. *Zeitschrift für Angewandte Linguistik* 69, 1–32. doi: 10.1515/zfal-2018-0015

Bybee, J. (2006). From usage to grammar: the mind's response to repetition. *Language* 82, 711–733. doi: 10.1353/lan.2006.0186

Bybee, J. (2013). "Usage-based theory and exemplar representations of constructions," in *The Oxford Handbook of Construction Grammar*, eds. T. Hoffmann and G. Trousdale (Oxford: Oxford University Press), 49–69.

Clark, H. H. (2016). Depicting as a method of communication. *Psychol. Rev.* 123, 324–347. doi: 10.1037/rev0000026

Council for Cultural Co-operation, Education Committee, and Modern Languages Division (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Croft, W., and Cruse, D. A. (2004). *Cognitive Linguistics.* Cambridge: Cambridge University Press.

Dancygier, B., and Vandelanotte, L. (2017). Internet memes as multimodal constructions. *Cogn. Linguist.* 28, 565–598. doi: 10.1515/cog-2017-0074

Diessel, H. (2023). *The Constructicon: Taxonomies and Networks*. Cambridge: Cambridge University Press.

Dingemanse, M., Blasi, D. E., Gary, L., Christiansen, M. H., and Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends Cogn. Sci.* 19, 603–615. doi: 10.1016/j.tics.2015.07.013

Divjak, D. (2019). *Frequency in Language: Memory, Attention and Learning.* Cambridge: Cambridge University Press.

Elvira-García, W. (2019). "Two constructions, one syntactic form: perceptual prosodic differences between elliptical and independent clauses in Spanish," in *Insubordination. Theoretical and Empirical Issues*, eds. K. Beijering, G. Kaltenböck, and M. S. Sansiñena (Berlin/Boston, MA: De Gruyter Mouton), 240–264.

Féry, C. (2017). *Intonation and Prosodic Structure.* Cambridge: Cambridge University Press.

Feyaerts, K., Brône, G., and Oben, B. (2017). "Multimodality in interaction," in *The Cambridge Handbook of Cognitive Linguistics*, ed. B. Dancygier (Cambridge: Cambridge University Press), 135–156.

Fillmore, C. J., Lee-Goldman, R. R., and Rhodes, R. (2012). "The FrameNet constructicon," in *Sign-Based Construction Grammar*, eds. H. C. Boas and I. A. Sag (Stanford: CSLI), 309–379.

Fry, D. B. (1955). Duration intensity as physical correlates of linguistic stress. *J. Acoust. Soc. Am.* 32, 765–769. doi: 10.1121/1.1908022

Fry, D. B. (1958). Experiments in the perception of stress. *Lang. Speech* 1, 126–152. doi: 10.1177/002383095800100207

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalizations in Language*. Oxford: Oxford University Press.

Gras, P., and Elvira-García, W. (2021). The role of intonation in construction grammar: on prosodic constructions. *J. Pragmat.* 180, 232–247. doi: 10.1016/j.pragma.2021.05.010

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hedberg, N., Sosa, J. M., and Fadden, L. (2003). "The intonation of contradictions in American English," in *Prosody and Pragmatics Conference*, 1–12. Available online at: https://www.sfu.ca/~hedberg/Preston_paper_text4.pdf (accession October 13, 2023).

Hiippala, T. (2017). The multimodality of digital longform journalism. *Digit. Journal.* 5, 420–442. doi: 10.1080/21670811.2016.1169197

Hilpert, M. (2019). *Construction Grammar and Its Application to English, 2nd Edn.* Edinburgh: Edinburgh University Press.

Hoffmann, T. (2017). Multimodal constructs – multimodal constructions? The role of constructions in the working memory. *Linguist. Vanguard* 3:20160042. doi: 10.1515/lingvan-2016-0042

Hoffmann, T. (2022). *Construction Grammar: The Structure of English.* Cambridge: Cambridge University Press.

Hoffmann, T., and Trousdale, G. (2013). *The Oxford Handbook of Construction Grammar.* Oxford: Oxford University Press.

Hsu, H. C., Brône, G., and Feyaerts, K. (2021). When gesture "takes over": speech-embedded nonverbal depictions in multimodal interaction. *Front. Psychol.* 11:552533. doi: 10.3389/fpsyg.2020.552533

Imai, M., and Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philos. Trans. Royal Soc. B* 369:20130298. doi: 10.1098/rstb.2013.0298

Imo, W., and Lanwer, J. P. (2020). *Prosodie und Konstruktionsgrammatik*. Berlin: De Gruyter.

Köhler, W. (1929). *Gestalt Psychology*. New York, NY: Liveright.

Kress, G. (2000). Multimodality: challenges to thinking about language. *TESOL Quarterly* 34, 337–340. doi: 10.2307/3587959

Kurumada, C., Brown, M., and Tanenhaus, M. (2012). Pragmatic interpretation of contrastive prosody: it looks like speech adaptation. *Proc. Ann. Meet. Cogn. Sci. Soc.* 34, 647–652.

Ladewig, S. (2020). Integrating Gestures: The Dimension of Multimodality in Cognitive Grammar. Berlin: De Gruyter.

Lanwer, J. P. (2020). "Appositive syntax oder appositive prosodie?" in *Prosodie und Konstruktionsgrammatik*, eds. W. Imo and J. P. Lanwer (Berlin: De Gruyter), 233–281.

Lehmann, C. (2023). "Multimodal markers of irony in televised discourse: a corpus-based approach," in *Multimodal Im/politeness: Signed, Spoken, Written*, eds. L. Brown, I. Hübscher, and A. H. Jucker (Amsterdam: Benjamins), 251–272.

Lehmann, C. (in press). "The prosody of irony is diverse and sometimes construction-specific," in *Interfaces of Phonetics*, ed. M. Schlechtweg (Berlin: De Gruyter).

Leiner, D. J. (2021). *SoSci Survey*. Available online at: https://www.soscisurvey.de (accessed June 18, 2023).

Lelandais, M., and Ferré, G. (2019). The verbal, vocal, and gestural expression of (in)dependency in two types of subordinate constructions. *J. Corpora Discour. Stud.* 2, 117–143. doi: 10.18573/jcads.4

Levinson, S. C. (2006). "Deixis," in *The Handbook of Pragmatics*, eds. L. R. Horn and G. Ward (Hoboken: Wiley), 97–121.

Levis, J. M., and Wichmann, A. (2015). "English intonation - form and meaning," in *The Handbook of English Pronunciation*, eds. M. Reed and J. M. Levis (Chichester: Wiley-Blackwell), 139–155.

Liberman, M., and Sag, I. A. (1974). "Prosodic form and discourse function," in *Papers from the Tenth Regional Meeting Chicago Linguistic Society*, eds. M. W. La Galy, R. A. Fox, and A. Bruck (Chicago, IL: Chicago Linguistic Society), 416–427.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *J. Acoust. Soc. Am.* 32, 451–454. doi: 10.1121/1.1908095

Long, J. A. (2022). *Jtools: Analysis and Presentation of Social Scientific Data*. Available online at: https://CRAN.R-project.org/package=jtools (accessed October 13, 2023).

Lüdecke, D. (2023). *sjPlot: Data Visualization for Statistics in Social Science.* R package version 2.8.15. Available online at: https://CRAN.R-project.org/package=sjPlot

Marandin, J.-M. (2006). *Contours as Constructions. Constructions Special Volume 1*. doi: 10.24338/cons-448

Neitsch, J., and Niebuhr, O. (2019). "Questions as prosodic configurations: how prosody and context shape the multiparametric acoustic nature of rhetorical questions in German," in *Proceedings of the 19th International Congress of Phonetic Sciences*, eds. S. Calhoun, P. Escudero, M. Tabain and P. Warren (Canberra, ACT: Australasian Speech Science and Technology Association), 2425–2429. Available online at: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS2019_Proceedings.pdf (accessed October 13, 2023).

Niebuhr, O. (2010). On the phonetics of intensifying emphasis in German. *Phonetica* 67, 170–198. doi: 10.1159/000321054

Niebuhr, O. (2019). "Pitch accents as multiparametric configurations of prosodic features – evidence from pitch-accent specific micro-rhythms in German," in *A Sound Approach to Language Matters - in Honor of Ocke-Schwen Bohn*, eds. A. M. Nyvad, M. Hejná, A. Hojen, A. B. Jespersen, and M. H. Sorensen (Aarhus: Aarhus University Press), 321–351.

Ningelgen, J., and Auer, P. (2017). Is there a multimodal construction based on non-deictic so in German? *Linguist. Vanguard* 3:20160051. doi: 10.1515/lingvan-2016-0051

Nolan, F. (2021). "Intonation," in *The Handbook of English Linguistics, 2nd Edn*, eds. B. Aarts, A. McMahon, and L. Hinrichs (Chichester: Wiley), 385–405.

Palan, S., and Schitter, C. (2018). Prolific Ac—a subject pool for online experiments. *J. Behav. Exp. Fin.* 17, 22–27. doi: 10.1016/j.jbef.2017.12.004

Perniss, P. (2018). Why we should study multimodal language. *Front. Psychol.* 9:e01109. doi: 10.3389/fpsyg.2018.01109

Põldvere, N., and Paradis, C. (2020). 'What and Then a little robot brings it to you?' The reactive *What-X* construction in spoken dialogue. *Engl. Lang. Linguist.* 24, 307–332. doi: 10.1017/S.1360674319000091

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.r-project.org/ (accessed June 23, 2022).

Sacks, H., Schegloff, E. A., and Jefferson, G. D. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.1353/lan.1974.0010

Sadat-Tehrani, N. (2010). An intonational construction. *Constructions* 3, 1–13. doi: 10.24338/cons-451

Schoonjans, S. (2018). *Modalpartikeln als Multimodale Konstruktionen: Eine Korpusbasierte Kookkurrenzanalyse von Modalpartikeln und Gestik Im Deutschen*. Berlin: De Gruyter.

Sidhu, D. M., Westbury, C., Hollis, G., and Pexman, P. M. (2021). Sound symbolism shapes the English language: the Maluma/takete effect in English nouns. *Psychon. Bullet. Rev.* 28, 1390–1398. doi: 10.3758/s13423-021-01883-3

Singer, N. 1. (2023). *Oxford English Dictionary*. Oxford: Oxford University Press.

Steen, F., and Turner, M. B. (2013). "Multimodal construction grammar," in *Language and the Creative Mind*, eds M. Borkent, B. Dancygier, and J. Hinnell (Stanford: CSLI), 255–274.

Stöckl, H. (2020). "Linguistic multimodality – multimodal linguistics: a state-of-the-art sketch," in *Multimodality: Disciplinary Thoughts and the Challenge of Diversity*, eds. J. Wildfeuer, J. Pflaeging, J. Bateman, O. Seizov, and C. I. Tseng (Berlin: DeGruyter), 41–68.

Tell, V. (2023). *In Oxford English Dictionary*. Oxford: Oxford University Press.

Tench, P. (1996). Intonation and the differentiation of syntactic patterns in English and German. *Int. J. Appl. Linguist.* 6, 223–256. doi: 10.1111/j.1473-4192.1996.tb00096.x

Uhrig, P. (2020). Multimodality in language and communication. *Zeitschrift für Anglistik und Amerikanistik* 68:4. doi: 10.1515/zaa-2020-2019

Uhrig, P. (2022). Hand gestures with verbs of throwing: collostructions, style and Metaphor. *Yearb. German Cogn. Linguist. Assoc.* 10, 99–120. doi: 10.1515/gcla-2022-0006

van Leeuwen, T. (2014). "Critical discourse analysis and multimodality," in *Contemporary Critical Discourse Studies*, eds. C. Hart and P. Cap (London: Bloomsbury), 281–296.

Vigliocco, G., Perniss, P., and Vinson, D. (2014). Language as a Multimodal phenomenon: implications for language learning, processing and evolution. *Philos. Trans. Royal Soc. Lond. Ser. B Biol. Sci.* 369:20130292. doi: 10.1098/rstb.2013.0292

Ward, N. G. (2019). *The Prosodic Patterns of English Conversation*. Cambridge: Cambridge University Press.

Wells, J. C. (2006). *English Intonation: An Introduction*. Cambridge: Cambridge University Press.

Wichmann, A., and Blakemore, D. (2006). The prosody-pragmatics interface. *J. Pragmat.* 38, 1537–1541. doi: 10.1016/j.pragma.2006.02.009

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., et al. (2023). *Ggplot2, Create Elegant Data Visualisations Using the Grammar of Graphics*. Available online at: https://CRAN.R-project.org/package=ggplot2 (accessed November 13, 2023).

Ziem, A. (2017). Do we really need a multimodal construction grammar? *Linguist. Vanguard* 3:20160095. doi: 10.1515/lingvan-2016-0095

Zima, E. (2017). On the multimodality of [all the way from X PREP Y]. *Linguist. Vanguard* 3:20160055. doi: 10.1515/lingvan-2016-0055

Zima, E., and Bergs, A. (2017). Towards a multimodal construction grammar. *Linguist. Vanguard* 3 :20161006. doi: 10.1515/lingvan-2016-1006