



OPEN ACCESS

EDITED BY
Kate Scott,
Kingston University, United Kingdom

REVIEWED BY
Alison Hall,
De Montfort University, United Kingdom
Caroline Jagoe,
Trinity College Dublin, Ireland

*CORRESPONDENCE
Pauline Madella
✉ pauline.madella@beds.ac.uk

RECEIVED 07 March 2023
ACCEPTED 30 August 2023
PUBLISHED 29 September 2023

CITATION
Madella P (2023) Relevance and multimodal
prosody: implications for L2 teaching and
learning. *Front. Commun.* 8:1181805.
doi: 10.3389/fcomm.2023.1181805

COPYRIGHT
© 2023 Madella. This is an open-access article
distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Relevance and multimodal prosody: implications for L2 teaching and learning

Pauline Madella*

School of Education and English, University of Bedfordshire, Bedford, United Kingdom

In this paper, I build on Scott's relevance-theoretic account of contrastive stress. Contrastive stress works as an extra cue to ostension in altering the salience of a particular constituent in an utterance and, as a result, the salience of one particular interpretation of that utterance. I draw on Scott's argument that contrastive stress does not encode procedural meaning. Contrastive stress is *unpredictable* and, as such, it is in confounding the hearer's expectations that it draws his attention to the accented word and prompt his search for different interpretive effects. I argue that contrastive stress is interpreted purely inferentially precisely because it is one of many pointing devices. It is to be interpreted by virtue of its interaction with other paralinguistic behaviors, all of which being different aspects of the same ostensive act of communication. This leads me to focus on the gestural nature of contrastive stress working as an act of pointing, which, as an ostensive communicative behavior, conveys that *if you look over there, you'll know what I mean*. Finally, I present the implications of analyzing contrastive stress in its multimodal context—as prosodic pointing—for the teaching and learning of L2 prosodic pragmatics and the development of interpretive abilities in the L2 hearer's mind.

KEYWORDS

relevance theory, contrastive stress, ostension, multimodal prosody, prosodic pointing, prosodic pragmatics, L2 pragmatics instruction

1. Introduction

“It is not what you said, it's *how* you said it!” (Culpeper, 2011). In English, the prosodic contours of an utterance are central in the conveyance of speaker meaning. In this paper, I focus on one most conspicuous prosodic pattern: “contrastive stress” (Sperber and Wilson, 1986/1995; Scott, 2021). English makes extensive use of contrastive stress and co-speech visual information, which together “enhances linguistic input, distorts it, or replaces it, and sometimes even contradicts it” (Rost, 2016, p. 42). With its extra “oomph”, contrastive stress draws the hearer's attention to one particular constituent of the utterance, often to result in contrastive reading. The syllable that carries the stress is signaled in upper case:

(1) SHE's always been the breadwinner.

While contrastive stress is ubiquitous in English, it is more or less accessible across languages (Ladd, 1996), but it remains a universal highlighting device: a vocal correlate of a pointing gesture (Sperber and Wilson, 1986/1995; Scott, 2021). This bears implications for L2 prosody and pragmatics development and pedagogy.

I begin in Sections 2 and 3 by building on Scott (2021) relevance-theoretic account of contrastive stress and further supporting her argument that contrastive stress is interpreted purely inferentially. In Section 4, I argue that this is largely due to contrastive stress being interpreted by virtue of its interaction with co-pointing behaviors and other “gestural accompaniments” (Jones, 1956), in its multimodal context. Contrastive stress is a special behavior because (1) it is the most conspicuous example

of multimodal prosody; one that English makes extensive use of, and (2) it is a vocal correlate of a pointing gesture. In Section 5.1, I focus on pointing as a “special” behavior, thereby bringing further evidence of why contrastive stress is special as a multimodal prosodic pattern par excellence. In Section 5.2, I demonstrate the pedagogical implications of my account of contrastive stress as prosodic pointing in the context of fine-tuning L2 hearer’s relevance mechanisms and understanding the pragmatics of L2 prosody.

2. Contrastive stress in English

In spoken English, prosodic patterns can be intentionally used to convey pragmatic meaning or “speaker meaning” (Wilson and Wharton, 2006). One such prosodic phenomenon par excellence is so-called “contrastive stress” (Sperber and Wilson, 1986/1995; Scott, 2017a,b, 2021). In English, what is commonly referred to as contrastive stress, although the terminology may vary in the literature, e.g., “prosodic contrastive focus” (Dohen et al., 2007), “prosodic pointing” (Loevenbruck et al., 2009), “contrastive focus” (Wells, 2006), “contrastive accent” (Bolinger, 1961), “nuclear heavy stress” (Haugen, 1949), is the use of marked tonicity, as opposed to unmarked tonicity. English is an intonation language, or pitch accent language (Wells, 2006). This means that there is a general tendency in English for the main pitch accent or “nucleus” to fall on the stressed syllable of the final content word of an intonation phrase (IP), as in (2):

(2) I’d love a COffee.

(2) is a case of unmarked tonicity, or neutral nucleus placement, what Chomsky and Halle (1968) would refer to as “normal stress” and describe as predictable. If the final content word repeats information, the nuclear accent will be shifted to highlight the last new piece of information, as in (3b). If the last content word in the IP highlights new, contrastive information, as in (3c), it will be accented:

- (3a) Would you like a COffee?
 (3b) I’d LOVE a coffee!
 (3c) I’d love a TEA!

In (3c), although the stress does fall on the last content word in the IP, i.e., tea, it is a case of marked tonicity, as it serves a contrastive function. For the nuclear accent to result in contrastive reading in (3c), some other unexpected element(s) would be added, such as a change in tempo, loudness, and duration. Stress is generally understood as “greater auditory prominence” (Katamba, 1989, p. 221–242); it is realized with “greater articulatory care” (Gussenhoven, 2004, p. 15). Contrastive stress is described as the most conspicuous accent of all (Bolinger, 1961, p. 83). Its extra oomph is produced by conveying “acoustic salience” through “increased intensity and duration” (Ladd, 1996, p. 58). “Loudness” is indeed presented as one of its distinctive traits (Bloomfield, 1933; Jones, 1956; Katamba, 1989; Wells, 2006). In marked tonicity cases, the nuclear accent can fall on “virtually any word which the speaker chooses to highlight” (Katamba, 1989, p. 242). This echoes Bolinger’s (1961, p. 96) argument that “one cannot predict with precision when, where, and how the shift will occur”, making the

location of the nucleus highly unpredictable, unless we are mind-readers (Bolinger, 1972). Consider how movement of the nucleus placement in the below utterances (4b–8b) results in the speaker producing different realizations (Clark, 2013) of one same sentence:

- 4(a) Is this the play you have been looking for?
 4(b) THAT is the play I have been looking for.
 5(a) Is this the play you have been looking for?
 5(b) That IS the play I have been looking for.
 6(a) Is this the book you have been looking for?
 6(b) That is the PLAY I have been looking for.
 7(a) Is this the play Gem’ has been looking for?
 7(b) That is the play I have been looking for.
 In (7b), the nuclear placement is signaled in bold.
 8(a) Is this the play you have been looking for?
 8(b) That is the play I WAS looking for.

In 4b–8b, the nuclear accent falls on an element of the utterance that is not typically accented yet, in so doing, reflects the speaker’s intention to produce meaningful effects. The accenting of a contrasting element draws the hearer’s attention and guides him in working out the speaker’s intended meaning. For example, in (7b), emphatic stress falls on “I” as opposed to Gem’ and thereby prompts the hearer to look for extra meaningful effects and infer that it is her, not Gem’, that had been looking for the play. The above examples (4b–8b) show that contrastive stress is used to draw attention to a constituent that is made to stand out for the hearer to believe that it bears some relevance to him and is worth processing.

In so-called marked tonicity, stress *per se* does not bear contrastive meaning (Scott, 2021). Scott’s account resonates with Bolinger’s (1961, p. 84) point that contrast is not a property of the accent itself but rather one of its functions being to “MEAN contrast”. By using contrastive stress, the speaker only guides, re-focuses, or re-directs the hearer’s attention, which results in a contrast. Dohen et al. (2007, p. 221) note that what they call “prosodic contrastive focus” is used to “emphasize a word or group of words in an utterance as opposed to another”. Thus, contrastive stress necessarily results in a contrast between the focused object and what has been deliberately left unaccented or deaccented. Not only does Scott (2021, p. 39) agree in arguing that contrastive stress does not encode contrastive meaning, but she goes further and argues that, in fact, it does not encode anything. In other words, the interpretation of contrastive stress is done purely inferentially. In so doing, Scott’s relevance-theoretic account of how contrastive stress is interpreted offers further insights into the nature of the inferential processes at play when processing and interpreting contrastive stress.

3. The relevance of contrastive stress

Sperber and Wilson’s relevance theory goes along with Grice’s (1967, p. 37) idea that “the very act of communicating creates expectations which it then exploits”. As such, an act of communication conveys to the hearer that paying attention to it will be worth their while. This is the basis for the Communicative

Principle of Relevance (Sperber and Wilson, 1986/1995, p. 260), defined in (9).

- (9) Communicative Principle of Relevance: Every act of ostensive communication communicates a presumption of its own optimal relevance.

Knowing how a hearer is likely to respond, the speaker can easily manipulate the effort to which the hearer is put and manipulate his expectations so as to trigger his search for effects which justify that effort (Sperber and Wilson, 1986/1995; Scott, 2017a,b, 2021). Prosodic patterns are used by the speaker so as to trigger the hearer's search for relevance and his expectation of positive cognitive effects:

A communicator who wants some prosodic feature of her utterance to be understood as contributing to her meaning should therefore do her best to make it salient enough, and rich enough in effects, to be picked out by the relevance-theoretic comprehension procedure and help make the utterance relevant in the expected way (Wilson and Wharton, 2006, p. 442).

As an extra cue to ostension (Scott, 2017a,b, 2021), an unexpected prosodic feature, such as contrastive stress, comes with the presumption that it is salient enough and rich enough in effects to be worth attending to and processing. Contrastive stress is salient not just because of the nuclear accent itself but due to the unexpectedness of the prosodic placement:

Any departure from neutral (or "expected") prosody would increase the hearer's phonological processing effort but would thereby encourage him to look for extra (or different) effects (Wilson and Wharton, 2006, p. 448).

In operating as an extra cue to ostension, contrastive stress comes with the presumption of its optimal relevance. It draws the hearer's attention to what he would have otherwise ignored, and focuses it on her intentions. This entails that the hearer is put to more effort only to raise his expectations of more or different cognitive effects (Scott, 2017a,b, 2021). Thus, contrastive stress does not lead to a "quick and cheap" inference (Tomlinson and Bott, 2013, p. 3569). It primarily *re-focuses* the hearer's attention, which causes it to be *effort-ful*. As House (2006, p. 1547) notes, "assigning salience orients the hearer to update her cognitive environment in a particular way". The updating of his cognitive environment or re-focusing of his attention necessarily involves extra processing effort on his part, which, concomitantly, raises the addressee's expectations of extra or different effects on the account that the speaker must have good reasons for re-orienting him in a particular way. Wilson and Carston (2019, p. 4) address precisely this point:

In language use, departures from expected syntax, wording or prosody [...] provide possible cues to ostension, focussing attention on particular aspects of the ostensive act and encouraging a search for additional interpretive effects.

As the most conspicuous accent of all, contrastive stress naturally stands out. It results in contrastive reading; however, it

does not *encode* contrast. As Scott (2021, p. 39) argues, contrastive stress is purely inferentially interpreted. As she explains, it is the disconfirmation of the addressee's expectations that triggers his search for different cognitive effects. In "confounding" the addressee's expectations, contrastive stress invites the hearer to follow a contrastive inferential route. Another point that Scott (2021) puts forward to support her argument is that contrastive stress does not activate the same procedure each and every time it is used, and so it cannot be said to encode procedural meaning. As Wilson (2016, p. 17) notes, it "merely point[s] the addressee in the right direction rather than providing a full concept as a starting point for inference". In other words, contrastive stress does not provide conceptually encoded content in the way that content words do, nor does it provide the addressee with a specific and systematic procedural instruction for him to follow (Fretheim, 2002) in the way that reference assignment does. Unexpected prosodic placement can be said, however, to have an impact on what Sax (2011, p. 378) names "procedures of comprehension", but it does not encode procedural constraints. Contrastive stress is unpredictable (Bolinger, 1972) in that it is a reflection of the speaker's choices as to what part of the utterance should be rendered more salient on the basis of the meaning she is intending to convey on that occasion and the inferential route the hearer needs to follow to arrive at the speaker's intended interpretation.

4. From contrastive stress to prosodic pointing

In this paper, I build on Scott (2021) relevance-theoretic account of contrastive stress. I draw on her argument that contrastive stress does not bear contrastive meaning nor encode procedural meaning. I support her account by suggesting that contrastive stress cannot be said to encode this or that procedural instruction precisely because it is interpretable by virtue of its interaction with co-pointing devices and other "gestural accompaniments" (Jones, 1956), provided that these are available to the hearer. In face-to-face communication, utterances generally are composites of a range of different behaviors, all of which being integral parts of the ostensive act of communication. As Ladd (1996, p. 40) notes, it is difficult to "unravel prosody from its paralinguistic context". As Wharton (2016, p. 5) also points out:

The parallels are so strong that a single, homogeneous account of these para-/non-linguistic behaviors seems to be required, one that embraces the fact that they are, for the most part, closely interlinked.

Psychologist McNeill (1985, p. 350) also describes those concomitant paralinguistic elements as "parts of a single psychological structure". It follows that contrastive stress must be considered in its multimodal context. Contrastive stress is not just a prosodic phenomenon; it is a multimodal phenomenon par excellence. Along with its gestural counterparts, contrastive stress plays an active part in "catching someone's eye, touching them, pointing, showing them something" (Wilson and Carston, 2019, p. 34). Contrastive stress is special precisely because it is probably the best illustration of *multimodal prosody*. Intonation in general, and contrastive stress specifically, is typically produced and interpreted

together with visual cues that play a crucial part in how prosodic patterns—in particular, unexpected prosodic patterns—are to be interpreted, as (10) illustrates:

(10) I did not know SHE was coming.

In (10), the words themselves come short of conveying the speaker's full intended meaning. The accenting of “she” is only one aspect of a larger gestural act of communication. To reach a hypothesis about the speaker's meaning, the addressee will likely process and incorporate the speaker's eye-, chin-, and head-pointing toward “she”, her frown and a face and tone of voice that show disapproval or discontentment, all contributing to revealing the attitude of the speaker and how the words are to be interpreted. As Stevick (1982, p. 163) expresses: “Nonverbal communication provides the surface on which the words are written and against which they must be interpreted”. In (10), the speaker communicates much more than what is said: The way it is said conveys that she is not particularly pleased to see that “she” is there, she's not friendly with “her”, etc. The co-pointing modalities coincide harmoniously with the vocally-conveyed highlighting of “she”, which is in line with research that has shown how “nods, hand gestures, and eye contact coincide very precisely with events in the spoken message” (Kendon, 1972; Ladd, 1996, p. 34). Beyond the decoding of the linguistic form, the para-linguistic features are salient enough to be read as relevant inputs to inferential processing, and it is on the basis of how they interact that the addressee is able to construct a hypothesis about the speaker's intended meaning by incorporating the pieces of the puzzle. These pieces or individual modalities may well be conceptual, but they will need to be adjusted in the process of interpreting the utterance through inferential work (Sperber and Wilson, 2015).

The very nature of utterances is complex, and what they communicate can be best described as “nebulous, contextually shaded and hard to pin down in conceptual terms” (Wharton, 2009, p. 146). As Madella and Wharton (2023) argue, it is by virtue of their interaction that the encoded concepts carried by individual modalities, such as a frown for disapproval, a nod for agreement, the vocal highlighting of a pronoun, eye-pointing are expected to be “adjusted or modulated in the course of the interpretation process” for the purpose of making one particular inference on that one particular occasion (Sperber and Wilson, 2015, p. 145). Scott (2021) makes a point that these modalities do not trigger the same procedural constraints each and every time they are used. It is indeed on the basis of its interaction with co-pointing modalities that its meaning is constructed, and so it is worked out purely inferentially. Sperber and Wilson's (1986/1995) theory of utterance interpretation involves going beyond the Gricean notion of speaker's meaning to accommodate the interpretation of vague and weaker communication. This is summarized by Wilson and Carston (2019, p. 34):

Relevance theorists set out from the start to look for a set of pragmatic principles and mechanisms that can deal with the full range of overtly intentional communicative acts: verbal and non-verbal, showing and telling, determinate and indeterminate, literal and figurative, propositional and non-propositional.

While non-verbal ostensive behaviors can be used to infer determinate, strong interpretations, they are often associated with vague, non-propositional, thus weaker communication, where it is difficult for the hearer to pinpoint one definite inference. In such cases, the speaker does not commit to one single interpretation but rather make “an array of roughly similar conclusions” available to the hearer (Wilson and Wharton, 2006, p. 1569). As contrastive stress is used as part of a wider range of composites all participating in the act of “showing”, pointing the hearer in the intended direction, its interpretation is bound to often be more of a “diffuse impression” (Wilson and Wharton, 2006, p. 1569). This is also more generally conveyed by cognitive scientists and psychologists Tomasello et al. (2007, p. 705) when they write that:

Pointing (...) does not convey a specific meaning in the manner of most conventionalized, symbolic gestures. Rather, pointing can convey an almost infinite variety of meanings by saying, in effect “If you look over there, you'll know what I mean”.

Considering contrastive stress in its multimodal context and, therefore, as purely inferentially interpreted, assumes a natural account of prosody. The pragmatic nature of prosody comes through from the intimate connections it entertains with gesture (Bolinger, 1983a,b,c). In other words, it is in its gestural dimension that the pragmatics of prosody shows; it is precisely where its pragmatic force lies and what makes its pragmatic nature visible. Gesture is what brings prosody and pragmatics together; it bridges the gap between prosody and pragmatics by reflecting the gestural dimension and pragmatic force of prosody (Madella, 2021). I follow a natural approach to prosody (Bolinger, 1983a,b,c), thereby presenting contrastive stress as a natural highlighting device and illustrating Bolinger's point that speech prosody is one part of a broader “gestural complex”. I focus on what I call *prosodic pointing*, or contrastive stress as one audio-visual construct. Adopting a natural approach to prosody, I contend that it is read the same way as gesture (Bolinger, 1983a,b,c) and treat prosody as gesture (Madella, 2021; Madella and Wharton, 2023). Thus, my perspective assumes a natural or universal approach to prosody, one that is in line with Bolinger's (1964) view of intonation as existing “around the edge of language”. The nature of prosody has been described as ranging from “natural” to purely linguistic (Wharton, 2009). Prosody has a dual nature (House, 2006), so prosodic meaning is best described as a matter of degree rather than an all-or-nothing distinction reflected in either a natural or grammatical account. Bolinger strongly favors the idea that although we may feel some aspects of intonation to be linguistic, those aspects retain a degree of naturalness and can easily be traced back to their natural origins:

Intonation... assists grammar—in some instances may be indispensable to it—but it is not ultimately grammatical... If here and there it has entered the realm of the arbitrary, it has taken the precaution of blazing a trail back to where it came from.

I, too, as far as contrastive stress is concerned, favor the view of prosody as a largely natural phenomenon, which belongs in the realm of pragmatics. This view contributes to our

understanding of contrastive stress as a multimodal phenomenon interpreted inferentially.

As a conclusion to his cross-linguistic study of accentuation variation, Ladd (1996, p. 167) argues against the idea of “some universal highlighting function” of prosody, which I disprove for reasons which will become apparent. The picture is indeed more complex as variability of accentuation is not consistent across languages. It is language specific (Sperber and Wilson, 1986/1995; Scott, 2021) and conditioned by the grammatical constraints of specific languages. In Norwegian, for instance, Fretheim (1998) explains that the word-accent system severely restricts the communicator’s intonational patterns. As illustrated in Section 2, English allows for flexible prosodic placement so long as it contributes to the speaker conveying her intended meaning. In other words, pragmatics can prevail over strict structural considerations. English is said to enjoy high pragmatically-motivated accentuation variability (Madella, 2021). This also suggests that pitch-marked prominence in English is more subject to unpredictability and a reflection of the speaker’s choices in comparison with languages that rely more heavily on structural constraints in their accent placement. Ladd’s (1996) shows that contrastive stress is less accessible cross-linguistically, while it is ubiquitous in English. However, this should not lead to the conclusion that the study fails to reveal the universal nature of contrastive stress. From a relevance-theoretic perspective, it shows that contrastive stress is more or less disruptive across languages (Wilson and Wharton, 2006; Wharton, 2009; Scott, 2017a). It will be less accessible to speakers of languages which do not place focal stress as freely as English does and make use of other, syntactic, constructions. French, for example, more typically uses cleft forms, as in (11a–c) below. The asterisk indicates that the utterance is ungrammatical or not typically used:

- (11a) C’est elle qui l’a fait. *It is her who did it.
 (11b) C’est ELLE qui l’a fait. *It is HER who did it.
 (11c) *ELLE l’a fait. SHE did it.

The syntactic extraction illustrated in (11a) is preferred over stressing “elle” to mark focus in French. This is not to say that contrastive stress in French is not at all possible, but cleft constructions are generally preferred. In (11b), both syntactic and prosodic contrastive focus (Dohen et al., 2007) are used. It is, however, used more sparsely and the accent is not quite equivalent to the intensity, duration, and loudness that characterize contrastive stress in English. That is due to the cleft construction contributing more heavily to the highlighting of the pronoun. French prosodic patterns do not allow for contrastive stress to be used as easily as it is used in English (Scott, 2021), and French has other preferred ways of conveying pointing, such as syntactic pointing. This is partly due to French being a non-intonation language. Similarly, in Spanish, the “a él” structure in (12a) would be preferred over the accenting of “lo” in (12b) (VanPatten, 2018):

- (12a) Bill lo conoce a él.
 (12b) *Bill LO conoce.
 (12c) Bill knows HIM.

The syntactic construction “a él” in (12a) will more likely be used to highlight “lo” (i.e., “him”). “A él” is the Spanish syntactic equivalent of prosodic stress on “him” in English (VanPatten, 2018). Another example of accentuation variability across languages is Italian (Ladd, 1996). Italian is known as a +rightmost language along with other languages, such as Spanish and Romanian. These languages resist *deaccenting*. In English, the accenting of an element that is typically unmarked necessarily entails that an element which would have been expected to carry the accent consequently becomes deaccented. The extensive use of contrastive stress contributes to deaccenting being an ordinary pattern in English, for example, in cases of repeated or given information. Semantic weight, semantic impoverishment, and semantic emptiness all are further conditions for deaccenting in English. However, +rightmost languages, like Italian, generally resist deaccenting of repeated material, empty content words, or last words:

- (13) I made a TRifle, but he HAtes desserts.

In the second intonation phrase of (13), “hates” rather than “desserts” would be accented, for “hates” is new information and so considered semantically richer as opposed to “desserts”, which is information already given by “Trifle”. In Italian, “desserts”, i.e., the +rightmost word, would typically be accented. It does not follow from Ladd’s study that contrastive stress cannot be regarded as a “natural” highlighting device across languages. What it does show is that contrastive stress is likely to be more or less disruptive across languages and, therefore, costlier and used more sparingly in those languages that rely more heavily on structural constraints (Sperber and Wilson, 1986/1995; Wilson and Wharton, 2006). As relevance-theorists argue, contrastive stress can be analyzed in terms of processing effort and cognitive effects. The process by which unexpected prosodic patterns put the hearer to extra processing effort and thus lead him to expect richer effects is universal (Wilson and Wharton, 2006; Scott, 2017a,b, 2021). The hearer is well aware that extra interpretive effects will likely offset the extra effort put in processing contrastive stress. In fact, contrastive stress is so routinely and ubiquitously used in English that it is expected to bear extra or different meaningful effects.

While Ladd (1996) demonstrates that the idea of intonation universals falls short in some way, the use of contrastive stress is often coupled with production of more universally recognized action, as demonstrated above. When Ladd (1996, p. 167) concludes that sentence accentuation is not “simply a matter of applying some universal highlighting gesture to individually informative words”, he is not far from claiming that a showing gesture or gestural highlighting would likely be more universal and would thus be less controversially recognized as natural. Bolinger’s description of a possibly pre-linguistic (almost biological) highlighting function of intonational contours used for the reading of speakers’ mental states and intentions has been controversial. His description, however, seems to suit an arguably less controversial pre-linguistic (and certainly biological) universal of human communication: *pointing*. According to Scott (2021, p. 37), contrastive stress, as an ostensive behavior, operates much like a pointing gesture. The speaker is “pointing to a

part of the utterance with her voice” (Scott, 2021). Imai (1998) describes prosody as a relevance indicator, some sort of natural “pointer” indicating where relevance is to be found. As Sperber and Wilson (1986/1995, p. 203) put it, “stress is a sort of vocal equivalent of pointing [...] a natural means of drawing attention to one particular constituent in an utterance”. Indeed, the deictic nature of contrastive stress makes it a very close equivalent to a pointing gesture. Scott’s (2017a) characterization of contrastive stress or vocal pointing provides further elements of support to why contrastive stress should be treated as one of many pointing modalities. Contrastive stress and pointing are (extra) cues to ostension, raising expectations and producing non-encoded meaning (Scott, 2017a):

- (1) They are both ostensive. They both prompts and guides the hearer’s inferential work. They focus attention and focus it on the speaker’s reasons for drawing his attention.
- (2) They both raise expectations of the ostensive stimuli’s optimal relevance. They both manipulate the hearer’s expectations in confounding them and thereby triggering his search for additional or different cognitive effects, which will justify and offset the extra processing effort required to retrieve the intended interpretation.
- (3) Both contrastive stress and a pointing gesture merely points the addressee in the intended direction without encoding anything. They are means of showing something and, in doing so, they guide the search for relevance.

Both contrastive stress and gestural pointing are driven by the same motivation (Madella, 2021); they are two aspects of the same process of utterance (Kendon, 1972, 1980). So based on Ladd’s conclusive remarks, contrastive stress *can* be seen as a natural universal highlighting device, one that is typically used as one of and along with many other pointing devices (Wilson, 2016). The argument for considering contrastive stress as a multimodal prosodic phenomenon appears even stronger when we look at pointing and why it is a special behavior.

5. Prosodic pointing is special: implications for L2 prosodic pragmatics

5.1. Pointing is special

Pointing is indisputably “special”, which makes contrastive stress a special multimodal phenomenon. For one thing, pointing lies at the root of human communication. It is ubiquitous and likely universal (Kita, 2003; Loevenbruck et al., 2008, 2009). A pointing gesture is typically performed “with the index finger and arm extended in the direction of the interesting object and with the other fingers curled inside the hand” (Butterworth, 2003, p. 9). Pointing in children is first expressed with both the eyes and the finger. It is then communicated via intonation, and finally with syntax. Ocular and manual forms of pointing are not the only way of expressing pointing through gesture, as example (10) has shown. Chin, eye gaze, and associated eyebrow motion could be added to the list, depending on which part of the world you are in. Lip-pointing, on the other hand, is not exactly common nor socially

recognized around Europe, but it is a widespread deictic gesture in Southeast Asia, the Americas, Africa and Oceania. A study of Lao speakers’ use of lip-pointing describes it as not only involving “protruding one or both lips, but also raising the head, sticking out the chin, lifting the eyebrows, among other things” (Enfield, 2001, pp. 185–191). In Māori gesture, eyebrow flashes are yet another specific form of pointing (Gruber et al., 2016).

Pointing is a “special” multimodal behavior in the brain as well. Loevenbruck et al.’s (2008; 2009) focuses on the more biological aspect of pointing and the cerebral domains that multimodal pointing recruits. They find that vocal and gestural pointing recruit similar cerebral domains; the two modalities are produced and perceived simultaneously (Loevenbruck et al., 2008, 2009). Loevenbruck et al.’s (2008; 2009) research on pointing is in line with the natural argument: If those pointing modalities entertain such intimate connections in the brain, it certainly shows that contrastive stress, as a paralinguistic, biological phenomenon should be discussed as one audio-visual construct. As they note, pointing, or a deictic behavior, is a “universal ability which orients the attention of another person so that an object/person/direction/event becomes the shared focus of attention” (Loevenbruck et al., 2008 p.1). The major role played by manual or indexical pointing in language development strongly suggests that “vocal pointing and pointing in other modalities may well be grounded in a common cerebral network” (Loevenbruck et al., 2008, p.1). This is also indicated by Hübscher and Prieto (2019), who describe gestural and prosodic development as “sister systems”, operating in parallel in the brain and jointly contributing to L1 socio-pragmatic development. Dohen et al. (2007) and Loevenbruck et al. (2008) suggest that the detection and perception of contrastive stress—what they call prosodic contrastive focus—relies on the reading of multimodal cues. Dohen et al. (2007) reported the results of Tong et al.’s (2005) study of the neural processes underlying the perception of contrastive stress as opposed to that of intonation for question and affirmation discrimination. Their results indicated that processing contrastive stress involves more diffused neural activity. Dohen et al. (2007) compared French participants’ perception of prosodic focus with that of syntactic pointing (used more typically in French). They found that processing syntactic pointing merely involved the frontal region of the brain, while processing prosodic contrastive focus—what I call contrastive stress—recruited *frontal and left parietal regions*. The left parietal regions are typically associated with other forms of pointing, such as gestural pointing. Perception and production of contrastive stress therefore seem to recruit multimodal activity. This was further supported by Dohen and Loevenbruck’s (2009) study on the interaction of audition and vision for the perception of prosodic contrastive focus. Their study (Dohen and Loevenbruck, 2009, p. 7) demonstrated that:

Even though the perception of prosodic focus is often considered as uniquely auditory, it is possible to perceive prosodic focus visually and the visual modality can enhance perception when prosodic auditory cues are degraded.

The above thus suggests that English speakers would recruit associative brain regions in their production and perception of contrastive stress. Dohen et al.’s work not only gives further motivation to look at contrastive stress as a gestural complex,

as one audio-visual construct, but it emphasizes the necessity to use multisensory information to detect contrastive stress in English and to consider the perception of contrastive stress as multimodal. The above neurological claims provide ample evidence that contrastive stress must be analyzed in its multimodal context, as a part of a broader audio-visual construct, which in turn offers further support to Scott's account according to which contrastive stress does not encode anything and is interpreted purely inferentially.

5.2. The relevance of prosodic pointing to L2 prosodic pragmatics development

As noted in Section 4, an analysis of contrastive stress as a multimodal prosodic phenomenon contributes to bridging the gap between prosody and pragmatics. Prosodic pointing—contrastive stress as one audio-visual construct—does a good job at illustrating the pragmatics of prosody or what Romero-Trillo (2012, 2016, 2019) calls *prosodic pragmatics*.¹ I have argued that the pragmatic force of prosody does not come from prosody alone. It lies in the gestural dimension of prosody and in the way that prosody naturally interacts with other paralinguistic communicative behaviors. I have argued and demonstrated (Madella and Romero-Trillo, 2019; Madella, 2021; Madella and Wharton, 2023) that analyzing contrastive stress as a multimodal construct bears important L2 pedagogical implications. Exposure to multimodal prosody generally, and prosodic pointing specifically, can be used toward fine-tuning L2 relevance mechanisms triggered by multimodal input to inferential processing. In other words, it enables L2 hearers to understand the speaker's non-verbal communicative behaviors as evidence of her intentions. It can therefore enhance L2 hearers'² ability and willingness to move beyond conceptual meaning and trust paralinguistic input in retrieving the speaker's intended interpretation. It was found that having access to prosodic pointing—after being exposed to contrastive stress alone—made L2 hearers appreciate the need to access multimodal input, for them to “remember more from visual information”, “understand more clearly because (they) can see the body-language”, and “see who speaks and their different faces” (Madella, 2021, p. 253, my amendment). Finally, it can develop the L2 hearer's alertness to the pragmatics of prosody and co-speech gesture and bodily accompaniments, which in turn contributes to the development of interpretive abilities in the L2 hearer. For instance, it was also found that the L2 hearer is more likely to understand the pragmatics of contrastive stress when it falls on “you” in the question “Would YOU like an apple?”, if he also has access to the speaker's gestural behavior, i.e., leaning forward and using an open-palm hand gesture showing that she is returning a question.

¹ Term used by Romero-Trillo (2012, 2016, 2019).

² The term “hearer” remains as it follows from the relevance-theoretic tradition. It does not imply that the L2 hearer does not listen attentively, intentionally, or purposely.

6. Conclusion

In this paper, I have built on Scott (2021) relevance-theoretic account of contrastive stress and further supported her argument that, as an unpredictable extra cue to ostension disconfirming the hearer's expectations, contrastive stress is interpreted purely inferentially. I put forward the argument that it is precisely because contrastive stress is typically interpreted in its multimodal context that its meaningful effects are to be interpreted purely inferentially by virtue of its interaction with co-speech gesture and co-pointing modalities. As an ostensive behavior, contrastive stress operates the same way as a pointing gesture does, and the gestural nature of contrastive stress justifies that we want to analyse it in relevance-theoretic terms as *prosodic pointing*. Analyzing contrastive stress as a multimodal phenomenon—as prosodic pointing—further supports Scott's argument that contrastive stress does not encode procedural meaning. It simply points the hearer in the intended direction, where evidence of the speaker's intentions is to be found. Finally, analyzing contrastive stress as a multimodal phenomenon bears implications for the development and instruction of L2 prosody and relevance mechanisms (Madella and Romero-Trillo, 2019; Madella, 2021; Madella and Wharton, 2023).

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

I am forever grateful to Tim Wharton for his continuing support. I would like to thank Frontiers and the editorial team for their trust and patience.

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 handling editor KS declared a past collaboration with the author PM.

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

- Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart & Winston.
- Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language* 37, 83–96. doi: 10.2307/411252
- Bolinger, D. (1964). Around the edge of language: intonation. *Harv. Educ. Rev.* 34, 282–296. doi: 10.17763/haer.34.2.4474051q78442216
- Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language* 48, 633–644. doi: 10.2307/412039
- Bolinger, D. (1983a). Where does intonation belong? *J. Semant.* 2, 101–120. doi: 10.1093/semant/2.2.101
- Bolinger, D. (1983b). Intonation and gesture. *Am. Speech* 58, 156–174. doi: 10.2307/455326
- Bolinger, D. (1983c). “The inherent iconicism of intonation,” in *Iconicity in Syntax*, ed J. Haiman. Amsterdam: John Benjamins, 97–109.
- Butterworth, G. (2003). Pointing is the royal road to language for babies, in *Pointing: Where Language, Culture, and Cognition Meet*, ed S. Kita. Mahwah, NJ: Erlbaum, 9–33.
- Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.
- Clark, B. (2013). *Relevance Theory*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139034104
- Culpeper, J. (2011). “It’s not what you said, it’s how you said it!”: prosody and impoliteness,” in *Discursive Approaches to Politeness. Mouton Series in Pragmatics [MSP]* (Berlin: Mouton de Gruyter), 57–83.
- Dohen, M., and Loevenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Lang. Speech* 52, 177–206. doi: 10.1177/0023830909103166
- Dohen, M., Loevenbruck, H., Callan, A., Callan, D., Baciú, M., Pichat, C., et al. (2007). ‘Multimodal Perception of Prosodic Contrastive Focus in French: A Preliminary fMRI Study’, *Summerschool: Cognitive and physical models of speech production, perception and perception-production interaction*. Autrans, France: Part II*: Brain and Speech.
- Enfield, N. J. (2001). ‘Lip-pointing’: a discussion of form and function with reference to data from Laos. *Gesture* 1, 185–212. doi: 10.1075/gest.1.2.06enf
- Fretheim, T. (1998). “Intonation and the procedural encoding of attributed thoughts: The case of Norwegian negative interrogatives,” in *Current Issues in Relevance Theory*, eds V. Rouchota and A. Jucker. Amsterdam: John Benjamins, 205–236.
- Fretheim, T. (2002). “Intonation as a constraint on inferential processing,” in *Proceedings of the Speech Prosody 2002 Conference, Aix-en-Provence, France*, eds B. Bell and I. Marlien, 59–64.
- Grice, P. H. (1967). “Logic and conversation” in *Unpublished MS from the William James Lectures*. Cambridge, Massachusetts: Harvard University.
- Gruber, J., King, J., Hay, J., and Johnston, L. (2016). The hands, head, and brow: a sociolinguistic study of Māori gesture. *Gesture* 15, 1–36. doi: 10.1075/gest.15.1.01gru
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Haugen, E. (1949). Phoneme or prosodeme? *Language* 25, 278–282. doi: 10.2307/410089
- House, J. (2006). Constructing a context with intonation. *J. Pragmat.* 38, 1542–1558. doi: 10.1016/j.pragma.2005.07.005
- Hübscher, I., and Prieto, P. (2019). Gestural and prosodic development act as sister systems and jointly pave the way for children’s sociopragmatic development. *Front. Psychol.* 10, 1259. doi: 10.3389/fpsyg.2019.01259
- Imai, K. (1998). “Intonation and Relevance,” in *Relevance Theory: Applications and Implications*, eds R. Carston and S. Uchida. Amsterdam: John Benjamins, 69–86.
- Jones, D. (1956). *An outline of English Phonetics*. Cambridge, UK: CUP.
- Katamba, F. (1989). *An Introduction to Phonology*. Harlow, Essex: Longman.
- Kendon, A. (1972). “Some relationships between body motion and speech,” in *Studies in Dyadic Communication*, eds A. Siegman and B. Pope. New York: Pergamon Press, 177–210.
- Kendon, A. (1980). “Gesticulation and speech: two aspects of the process of utterance,” in *The Relationship of Verbal and Nonverbal Communication*, ed M. R. Key. The Hague: Mouton and Co., 207–227.
- Kita, S. (2003) *Pointing: Where Language, Culture and Cognition Meet*. Mahwah, NJ: Erlbaum.
- Ladd, R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
- Loevenbruck, H., Dohen, M., and Vilain, C. (2008). From gestural pointing to vocal pointing in the brain. *Revue Française de Linguistique Appliquée* 8, 23–33. doi: 10.3917/rfla.132.0023
- Loevenbruck, H., Dohen, M., and Vilain, C. (2009). “Pointing is “special””, in *Some Aspects of Speech and the Brain*, eds S. Fuchs, H. Loevenbruck, D. Pape, and P. Perrier. Frankfurt am Main: Peter Lang, 211–258.
- Madella, P. (2021). *Prosodic Pointing: from Pragmatic Awareness to Pragmatic Competence in Chinese Hearers of L2 English*. Brighton and Hove: University of Brighton.
- Madella, P., and Romero-Trillo, J. (2019). Prosodic pointing in inferential comprehension: the application of relevance theory to L2 listening instruction. *Letrónica* 12, 1–17. doi: 10.15448/1984-4301.2019.4.34157
- Madella, P., and Wharton, T. (2023). “Non-verbal communication and context: multi-modality in interaction,” in *The Cambridge Handbook of Language and Context*, ed J. Romero-Trillo (Cambridge University Press), 419–435.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psych. Rev.* 92, 350–371. doi: 10.1037/0033-295X.92.3.350
- Romero-Trillo, J. (2012) *Pragmatics, Prosody and English Language Teaching*. Dordrecht: Springer.
- Romero-Trillo, J. (2016). “The pragmatics of prosody, the prosody of pragmatics [Plenary talk],” in *Séminaire de Recherche en Linguistique*. Geneva: University of Geneva, Switzerland.
- Romero-Trillo, J. (2019). Prosodic pragmatics and feedback in intercultural communication. *J. Pragmat.* 151, 91–102. doi: 10.1016/j.pragma.2019.02.018
- Rost, M. (2016). *Teaching and Researching Listening*. New York, NY: Routledge. doi: 10.4324/9781315732862
- Sax, D. (2011). “Sentence Stress and the Procedures of Comprehension,” in *Procedural Meaning: Problems and Perspectives (Current Research in the Semantics/Pragmatics Interface)*, eds V. Escandell-Vidal, M. Leonetti, and A. Ahern. Bingley: Emerald Group Publishing, 349–381.
- Scott, K. (2017a). *Ostension, Expectations and Non-Encoded Meaning [Presentation]*. Belfast, UK: International Pragmatics Association (IPra) Conference.
- Scott, K. (2017b). “Prosody, procedures and pragmatics,” in *Semantics and Pragmatics: Drawing a Line Series: Logic, Argumentation & Reasoning*, eds L. Depraetere and R. Salkie. Cham: Springer, 323–341.
- Scott, K. (2021). “Contrastive stress in English: meaning, expectations and ostension,” in *Pragmatics and Beyond New Ser. Beyond Meaning*, eds T. Wharton, L. D. Saussure, E. Ifantidou (Amsterdam: John Benjamins).
- Sperber, D., and Wilson, D. (1986/1995) *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Sperber, D., and Wilson, D. (2015). Beyond speaker’s meaning. *Croat. J. Philos.* XV, 117–149.
- Stevick, E. W. (1982). *Teaching and Learning Languages*. Cambridge: Cambridge University Press.
- Tomasello, M., Carpenter, M., and Liszkowski, U. (2007). A new look at infant pointing. *Child Dev.* 78, 705–722. doi: 10.1111/j.1467-8624.2007.01025.x
- Tomlinson, J. M., and Bott, L. (2013). “How intonation constrains pragmatic inferences,” in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth, 3569–3575.
- Tong, Y., Gandour, J., Talavage, T., Wong, D., Dziedzic, M., Xu, Y., et al. (2005). Neural circuitry underlying sentence-level linguistic prosody. *NeuroImage* 28, 417–428. doi: 10.1016/j.neuroimage.2005.06.002
- VanPatten, B. (2018). “The (risky) promise of instructed SLA [Presentation],” in *51st Annual Meeting of the British Association of Applied Linguistics (BAAL)*, York, UK.
- Wells, J. C. (2006). *English Intonation: An Introduction*. Cambridge: Cambridge University Press.
- Wharton, T. J. (2009). *Pragmatics and Non-Verbal Communication*. Cambridge: Cambridge University Press.
- Wharton, T. J. (2016). “Paralanguage,” in *The Routledge Handbook of Pragmatics Routledge Handbooks in Applied Linguistics*, eds A. Barron, G. Steen, and G. Yueguo (London: Routledge), 69–75. doi: 10.1515/ip-2017-0032
- Wilson, D. (2016). Reassessing the conceptual-procedural distinction. *Lingua* 175, 5–19. doi: 10.1016/j.lingua.2015.12.005
- Wilson, D., and Carston, R. (2019). Pragmatics and the challenge of ‘non-propositional’ effects. *J. Pragmat.* 145, 31–38. doi: 10.1016/j.pragma.2019.01.005
- Wilson, D., and Wharton, T. J. (2006). Relevance and prosody. *J. Pragmat.* 38, 1559–1579. doi: 10.1016/j.pragma.2005.04.012