

Laura Ruth-Hirrel* and Sherman Wilcox

Speech-gesture constructions in cognitive grammar: The case of beats and points

<https://doi.org/10.1515/cog-2017-0116>

Received October 12, 2017; revised March 31, 2018; accepted April 11, 2018

Abstract: The current study uses principles from Cognitive Grammar to better account for the symbolic integration of gesture and speech. Drawing on data collected from language use, we examine the use of two attention-directing strategies that are expressed through gesture, beats and pointing. It has been claimed that beats convey no semantic information. We propose that beat gestures are symbolic structures. It has also been noted that beats are often overlaid on other gestures. To date, however, no detailed explanation has been offered to account for the conceptual and phonological integration of beats with other co-expressed gestures. In this paper, we explore the integration of beats and pointing gestures as complex gestural expressions. We find that simple beat gestures, as well as beat gestures co-expressed with pointing gestures, are used to direct attention to meanings in speech that are associated with salient components of stancetaking acts. Our account further reveals a symbolic motivation for the apparent “superimposing” of beats onto pointing gestures. By closely examining actual usage events, we take an initial step toward demonstrating how the symbolic elements of both beats and points are integrated in multimodal constructions.

Keywords: gesture, cognitive grammar, multimodality, stancetaking, directing attention

1 Introduction

Not every word a speaker utters has equal status. Some meanings are more important to the communicative goals and intentions that a speaker has for a particular interaction. Languages typically offer a variety of expressive resources that function to call special attention to meanings that a speaker wants to

*Corresponding author: Laura Ruth-Hirrel, Department of Linguistics, California State University, Northridge, CA, 91330, USA, E-mail: lahirrel@gmail.com

Sherman Wilcox, Department of Linguistics, The University of New Mexico, Albuquerque, NM, 87131, USA, E-mail: wilcox@unm.edu

emphasize. For example, English speakers can direct attention to different aspects of meaning using pitch accents. In the contrastive focus construction *He didn't BUY the car from her. She GAVE it to him*, pitch accents emphasize a contrast between the negated proposition in the first clause and the affirmative proposition that follows. Morphosyntactic strategies are also available in many languages for attention-directing functions and can be used in combination with prosodic strategies. An example from English is the topic-focus construction in which the direct object is moved to a position typically reserved for the subject, giving further emphasis to a clause-level topic (e.g., *I enjoyed visiting the Colosseum but **the FORUM**, I liked the best.*). Most languages also have lexical inventories that are used to intensify or highlight particular words. English has a number of words that are used as degree modifiers to intensify or emphasize a property associated with a referent (e.g., ***very** smart, **so** sloppy, **completely** wrong, **ridiculously** funny*). These strategies reflect the importance of speakers' need to highlight or draw attention to different dimensions of meaning in language use.

In this paper we examine the use of two attention-directing strategies that are expressed using gesture: beats and pointing. Beat gestures have been formally characterized as “biphasic movements of the hands” (Biau et al. 2015). These movements typically involve a “simple flick of the hand or finger up and down, or back and forth” (McNeill 1992: 15); however, beats may also be performed using other body parts, such as the head or eyebrows (Krahmer and Swerts 2007). Certain phases in the movement of the beat closely interact with temporal structures of the accompanying speech. The apex of a beat movement, which is the “kinetic ‘goal’ of the stroke” (Loehr 2004: 89) or the “point of maximal extension” (Alexanderson et al. 2013; Leonard and Cummins 2011: 1459), temporally aligns with pitch accented syllables (Alexanderson et al. 2013). The temporal alignment of beat gestures with speech has been found to impact the perception of prominence at the word level. Listeners perceive words that are co-expressed with beats (including those beats performed non-manually) as having a higher degree of prominence than words that do not occur with beats (Krahmer and Swerts 2007). This effect is observed even when the timing of beats in relationship to speech is manipulated to be misaligned with pitch accents.

The function of beats has been described in various ways. Many researchers have claimed that beats have “no semantic content” (Alibali et al. 2001; Biau and Soto-Faraco 2013; Özçalışkan and Goldin-Meadow 2009), perhaps considering semantics to include only propositional meanings. Many, including some researchers who characterize beats as being without semantic content, have acknowledged that beats serve emphatic functions and are closely tied to

information structure (McNeill 1992; McNeill et al. 2015; Theune and Brandhorst 2010). McNeill et al. (2015) suggest that beats serve multiple functions related to the structuring of information and discourse status. They compare the different functions of beats to an “all-purpose highlighter” (2015: 274). For example, beats may be used to signal a shift in the discourse mode, such as from a metanarrative (commentative) mode to a narrative (descriptive) mode. Other beat gestures, particularly those described as being temporally “superimposed” on other types of gestures, may emphasize that “the gesture (and concomitant speech) have significance beyond itself, in the larger context” (2015: 275). This point is important because co-speech gestures have long been analyzed as holistic structures that do not combine with other gestures to form more complex constructions (McNeill 1992: 21).

Research examining how beats are processed is suggestive regarding their role in directing attention (Biau and Soto-Faraco 2015; Holle et al. 2012). Biau and Soto-Faraco (2013) measured brain activity (i.e., event-related potentials) while participants observed videos of language use that included beats. They found that beats influenced the processing of speech, acting as attention attractors to the words and meanings with which they were co-expressed. Research on the relationship between beat gestures, prosody, and speech perception provides evidence for the integration of beats and speech at the time of conceptualization. However, missing from research on beat gestures is an exploration into the conceptual characteristics of this multimodal integration.

Like beat gestures, pointing also serves an attention-directing function. Pointing has been extensively studied by gesture researchers, as well as spoken and signed language linguists. Kita (2003) defines pointing as a communicative body movement that projects a vector from a body part indicating a certain direction, location, or object. Kendon (2010: 20) defines pointing both formally and conceptually, stating that pointing is “a form of visible action that serves to establish that something external to the speaker or his discourse, is the referent or topic of it.” Pointing, he says, is often accomplished with the hand, “but it may also be done with the head, chin, lips, or eyes, by a jerk of the elbow, a movement of the foot, by a bend of the torso, a flexion of a shoulder”.

Directing attention is recognized as a common function of pointing. Goodwin (2003) notes that pointing takes place in a communicative situation that contains at least two participants, one of whom performs a pointing gesture to establish a particular space as a shared focus for cognition and action. Clark (2003) distinguishes between pointing and placing, the former being a type of directing-to, and the latter placing-for, as two forms of indicating. In pointing, speakers *direct* their addressees’ attention *to* the objects (concrete or conceptual)

they are indicating, whereas in placing, speakers try to *place* the object they are indicating so that it falls within the addressees' focus of attention.

Langacker (2016b) finds that pointing serves a referential grounding semantic function. In his analysis pointing has a *directive force*, instructing the hearer to follow the direction of the pointing gesture so that both the speaker and the hearer focus their attention on the same entity, which is the referent of the gesture. Wilcox and Occhino (2016) propose that pointing in signed languages is a construction comprised of two component symbolic structures: a *pointing device*, such as a hand, serving to direct attention (visual or conceptual) to a second component structure, which they call a *Place*.

The literature on pointing reveals two distinct but ultimately related issues. First, while gesture researchers often note that pointing consists of two elements – e.g., a body *movement* indicating a *location*, or a visible *action* establishing a *referent* – they do not offer an explicit analysis of the symbolically complex nature of pointing. Second, many researchers agree that pointing serves to direct the attention of the addressee so that the speaker and addressee focus their joint attention on the same referent. We will suggest that a symbolic account of the composite nature of pointing helps to reveal more specifically how it serves to direct attention.

The current study uses principles from Cognitive Grammar (CG) (Langacker 1987, 2008) and multimodal construction grammar (Stickles 2016; Kok and Cienki 2015) to account for the symbolic integration of gesture and speech. Drawing on data collected from language use, we first analyze the use of beat gestures in multimodal constructions. We define constructions as form-meaning mappings which vary along a continuum of conventionality. This study agrees with researchers who have suggested that beats interact with information structure. However, in taking a cognitive approach, we reject claims that beats do not express semantic content, as information structure is understood to be a part of meaning. Instead we provide a detailed account of how the schematic meaning of beats, which we describe as emphasis, is expressed through the participation of beats in specific constructions. As previously mentioned, it has been claimed that beats are often overlaid on other gesture types (McNeill et al. 2015). To date, however, no detailed explanation has been offered to account for the conceptual and phonological integration of beats with other co-expressed gestures. In this paper we explore the integration of beats and pointing gestures in complex gestural expressions. We find that beat gestures alone, as well as beat gestures co-expressed with pointing gestures, are used to direct attention to meanings in speech that are associated with the key components of stancetaking acts. While simple beats and beat-point constructions both interact with the expression of

stancetaking, we will show that beat-point constructions serve different functions than simple beats.

Through this analysis, we challenge the view still present in the literature that gestures do not combine with other gestures to form complex constructions. In fact, our account shows a symbolic motivation for the apparent “superimposing” of beats onto pointing gestures. By closely examining actual usage events, we take an initial step toward demonstrating how the symbolic elements of both beats and points are specifically integrated in complex multimodal constructions.

2 Cognitive grammar

The central claim of CG is that only three structures are posited (Langacker 1987): semantic, phonological, and symbolic. Semantic structures are conceptualizations exploited for linguistic purposes. Phonological structures include sounds, gestures, and orthographic representations; an essential feature of phonological structures is that they are able to be perceived. Symbolic structures form an associative link between phonological and semantic structures, such that one is able to evoke the other.

Symbolic structures are abstracted from discourse. CG views discourse as the ongoing succession of *usage events*, actual instances of language use (Langacker 2001). By extracting commonalities across usage events, speakers develop *schemas*, superordinate concepts which specify what is common to several, or many, more specific concepts (Tuggy 2007). The more specified structures are called *elaborations* or *instantiations* of the schema. Schematicity is a relative notion: a concept is schematic in relation to a more specific concept, and an elaboration is more specific relative to a higher-level schema. Schematization applies to both the phonological and the semantic poles of symbolic structures.

Discourse takes place within a shared *ground*, which consists of the speech event, the speaker (S) and hearer (H), their interaction, their conception of reality, and the time and place of the speech event.¹ Discourse also takes place within a “current discourse space” (CDS), which is “the mental space comprising those elements and relations construed as being shared by the speaker and hearer as the basis for communication at a given moment in the flow of discourse” (Langacker 2001: 144). One goal of discourse is that the speaker and hearer focus their attention on the same conceived entity within

¹ When we refer to a speech event, we also include gesture. Thus, the ground also includes gesture space.

this shared discourse space. The symbolic resources available to speakers permit them to manage their limited attentional and conceptual “field of view,” akin to the visual field of visual perception. As Langacker (2001: 145) explains, “Metaphorically, it is as if we are ‘looking at’ the world through a window, or *viewing frame*. The *immediate scope* of our conception at any one moment is limited to what appears in this frame, and the *focus* of attention – what an expression *profiles* (i.e., designates) – is included in that scope” (Figure 1).

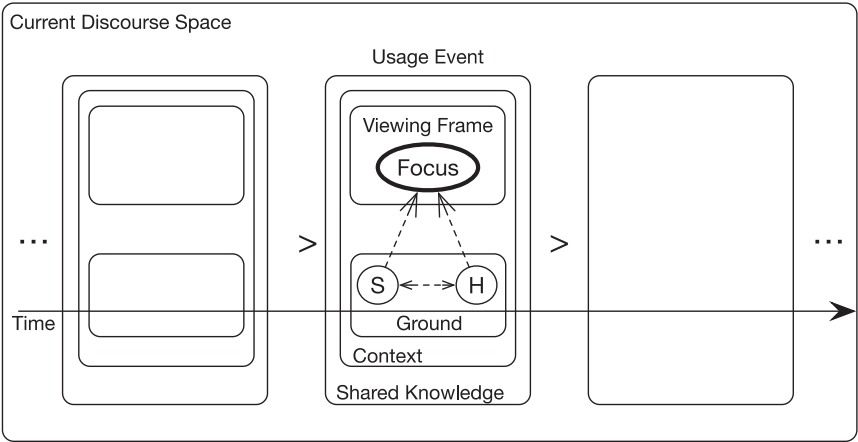


Figure 1: Usage event and viewing frame (from Langacker 2001: 145).

As shown in Figure 2, symbolic structures incorporate multiple *channels* (Langacker 2001). The semantic pole consists of several conceptualization channels, including speech management, information structure, and objective

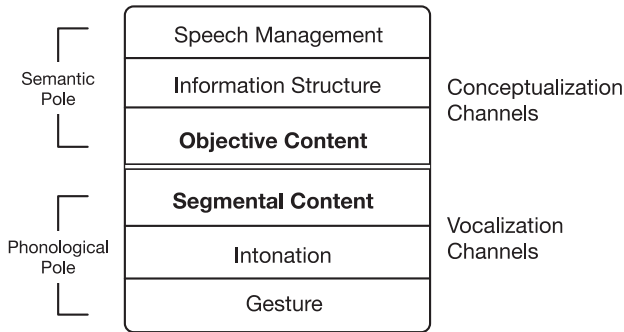


Figure 2: Conceptualization and vocalization channels (modified from Langacker 2001).

content. Speech management includes such functions as holding the floor and turn taking. Information structure includes emphasis, discourse topic, and given vs. new information. Objective content is a core channel; it is the conceptualization of the situation being described by a linguistic expression. The phonological pole consists of several vocalization channels. The core vocalization channel for speech is segmental content. Other channels include intonation and gesture.

Symbolic structures combine with other symbolic structures to form complex symbolic assemblies. As Langacker (2008: 161) notes, “Most of the expressions we employ are symbolically complex, to some degree analyzable into smaller symbolic elements. Grammar consists of the patterns for constructing such expressions. Accordingly, the expressions and the patterns are referred to as constructions. ... Constructions are symbolic assemblies”. This characterization suggests nothing about the status of a construction in terms of conventionality within a speech community. Symbolic assemblies (i.e., constructions) vary in the degree to which they are conventional. While individual component structures might be highly conventional, the integration of those components as a symbolic assembly might be more creative, although it is expected that at least some degree of conventionality must exist for communication to be successful. Likewise, a symbolic assembly might be entrenched for a particular speaker without being conventional across a community of speakers.

We claim and will offer evidence that in addition to constructions in the spoken modality, component symbolic structures that are vocalized in the gestural channel integrate to form more complex gestural constructions. Constructions also may consist of symbolic assemblies composed of speech and gesture. These latter are thus *multimodal constructions* consisting of the phonological and semantic integration of speech and gesture.

Both speech and gesture exhibit an asymmetry, which in CG is called *autonomy* and *dependence*. As Langacker (2016a: 10) explains,

An autonomous structure (A) has the potential to be manifested independently. A dependent structure (D) requires the support of an autonomous one for its full manifestation: ((A) D). It thus makes schematic reference to A as part of its internal structure. Being autonomous, A is usually more substantive than D, and by definition it has priority. So ... we can say that D elaborates A to form a higher-level structure AD. Because it incorporates A, this higher-level structure is normally autonomous as well and may in turn be elaborated. Langacker (2016a: 10)

Autonomy and dependence are features of both the phonological and the semantic poles of symbolic structures. A/D organization can be observed both within a particular channel and across channels of the same pole. For example,

internal to the spoken segmental content channel, vowels are autonomous in relation to consonants. A/D structure also is manifest across channels of the same pole. Intonation is dependent, making schematic reference to an autonomous carrier, the segmental content, for its full manifestation. Likewise, the information structure conceptualization channel is to some extent dependent, requiring information that is emphasized or serves as the target of discourse focus; this autonomous information is often provided by the objective content channel.

While the various channels correspond to qualitatively different modes of articulation/vocalization at the phonological pole and to different dimensions of meaning at the semantic pole, it would be erroneous to interpret them as sharply distinct from one another. Usage events come to us as complex symbolic assemblies that simultaneously incorporate multiple vocalization channels and evoke meanings across and within each of the conceptual channels. One benefit to referring to channels when talking about symbolic structures is that it emphasizes that the semantic pole includes more than strictly propositional meanings (see also Pisoni 1997). Higher-level grammatical constructions that express meanings associated with information structure and speech management are symbolic in their own right. All of these types of meanings are aspects of the semantic pole, they just emphasize different dimensions of it, as will be seen in the case of beats and pointing gestures.

3 The gestural channel

Within the CG framework, the term *gesture* identifies a vocalization channel, the phonological pole of a symbolic structure. Like the segmental content channel, the gestural channel also exhibits A/D asymmetry. In order to describe this aspect of the gesture channel, we use a classification first proposed by sign linguists (Battison 1978; Stokoe 1960) for describing the phonological structure of signed languages. Like signs, manual gestures consist of handshapes, locations, orientations, and movements. Handshapes are autonomous physical entities composed of material substance residing in space. In a specific gestural construction the hands occupy a location in space and an orientation (the direction in which the palm faces). Hand location and orientation are dependent properties requiring an autonomous hand for their manifestation. Movement is also a dependent property: a movement makes schematic reference to an autonomous entity (which moves). In gesture, the autonomous entity is typically instantiated by the hand(s).

We include one additional articulatory property, manner of movement, which is essential to our account of the phonological structure of beat gestures. The way in which a particular movement is produced is a dependent property, requiring a more autonomous property, which is provided by movement. Although a full treatment of what constitutes manner of movement is beyond the scope of this article, a simple example will demonstrate the basic notion. Consider the difference between tracing a linear path with the index finger, arm outstretched, moving downward (e.g., to depict a painted vertical line on a wall) in a slow and steady motion, and the gesture an orchestra conductor makes to give the opening downbeat of a piece of music. The downbeat is given with a distinct, rapid initial acceleration. This is a quality of motion we are calling manner of movement.²

Thus, the gestural channel is internally complex. Certain aspects of gestural form are autonomous and other aspects are dependent. These A/D aspects of gesture will play a significant role in the integration of beat and point gestures, as well as their integration with speech.

4 A symbolic analysis of multimodal assemblies

Co-speech gestures, we claim, are symbolic structures: they combine a phonological pole specified in the gestural channel, and a semantic pole specified in at least one of the conceptualization channels. Highly conventional gestures such as the emblems “V for victory” or the “thumb-up” gesture are phonologically specific gestures semantically specified in the objective content channel: they have specific phonological instantiations and specific semantic content. They thus function like, and in many cases can substitute for, spoken lexical items, which are phonologically and semantically specific as well. As we will see in the next section, gestures may also be specified in another conceptual channel.

We draw upon data collected from language use to support our claims that (1) beats are symbolic structures, (2) gestural symbolic structures integrate with other gestural symbolic structures to form complex gestural constructions, and (3) gestural constructions are integrated with speech to form even more complex multimodal constructions. In 4.1, we examine functions that simple beat

² A fuller treatment of manner of movement will require delving into the physics of motion, specifically the derivatives of location. The first derivative is velocity, or the rate of change of location; the second is acceleration (rate of change of velocity); the third is jerk (rate of change of acceleration). Our sense is that manner of movement (of path movement) involves acceleration and jerk. Other types of movements (e.g., cyclic) will require further analysis.

component structures serve in multimodal composites. In 4.2, we turn to more complex gestural constructions that include the co-expression of beat and point gestures. We explore the ways in which beat-point composites in the gestural channel are symbolically integrated with speech in multimodal constructions. We also examine symbolic motivations for the integration of beats and points in complex gestural constructions. We find that across the diversity in the specific contexts and constructions with which beat and point gestures are, both types of gestures play important roles in stancetaking acts and in the directing of attention.

4.1 Analysis 1: beats in multimodal constructions

We claim that beat gestures are symbolic structures. Figure 3 illustrates the schematic phonological and semantic structure of beat gestures. Beats are expressed phonologically as manner of movement. Semantically, beats are specified in the information structure channel. A primary semantic function of beats is to emphasize or highlight some material external to the beat, typically expressed in the spoken modality.

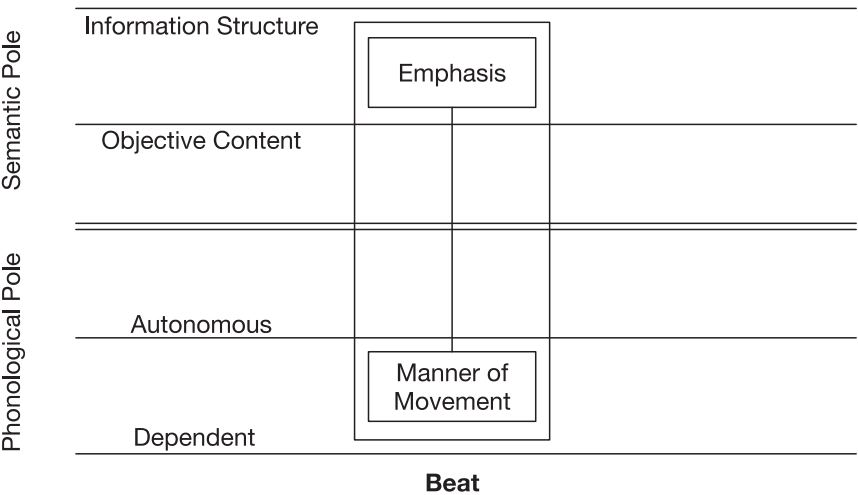


Figure 3: Beat symbolic structure.

A significant property of beats is that they are phonologically and conceptually dependent structures, requiring autonomous structures for their instantiation.

Phonologically, beats require an autonomous gesture *carrier* for their articulatory elaboration (indicated by the bold arrow in Figure 4). This is canonically specified by a handshape or, more accurately, by the movement of a handshape, since manner of movement is dependent on movement. It is possible for the movement of any more substantive and autonomous structure to serve as the phonological carrier for a beat. Beats are also phonologically dependent on symbolic structures expressed in spoken language. Beats align with and are dependent on pitch accents, which are specified in the intonation channel. Pitch accents are also dependent, making schematic reference to segmental content.

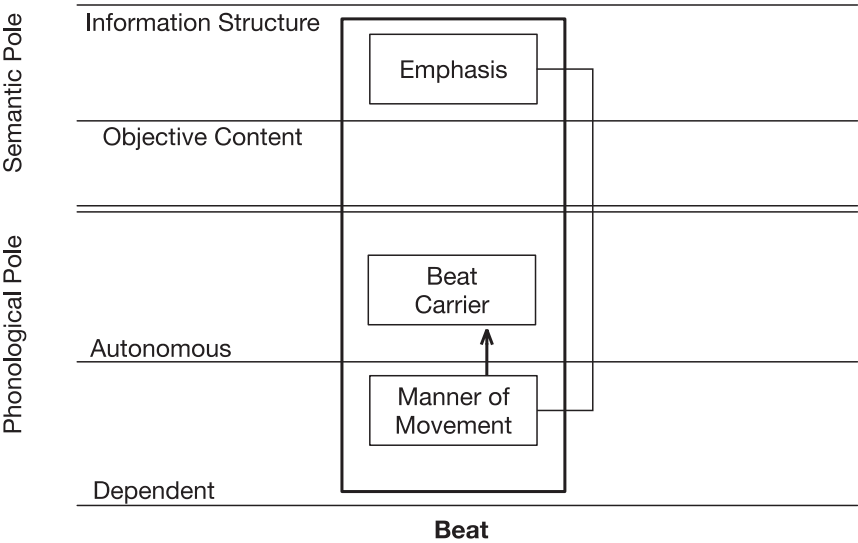


Figure 4: Beat carrier.

Semantically, beats are dependent structures making reference to some autonomous content, the information that is emphasized or highlighted. The emphasized information is typically specified in the objective content channel of the accompanying speech. As we will see in our data analysis, because speech tends to be expressed as complex constructions that integrate objective content with information structure, beats exhibit high degrees of semantic complexity.

Example (1), taken from a popular U.S. television talk show, shows a multi-modal construction that includes a series of simple beats.³

³ Methods of data collection and coding are described in Appendix 1, and coding convention in Appendix 2.

1. I have to **say** Molly **Shannon** in **this movie** gives the most **unsentimental brave performance**

At this point in the interaction, after host Stephen Colbert says, “So let’s talk about the movie”, actor Bradley Whitford answers questions about the plot of a film in which he co-starred alongside actor Molly Shannon. Then, in (1), Whitford performs a stancetaking act in which he subjectively evaluates his co-star’s performance in the film. During the expression of the evaluative stance act in the spoken utterance, Whitford uses several iterations of beat gestures. The beats in this example are expressed as a stressed manner of movement on a recurring downward movement of the speaker’s right hand, which is holding a pair of glasses (see Figure 5).⁴ In this example the beat carrier is not symbolic; rather it simply serves the instrumental function of holding glasses.



Figure 5: First beat produced in example (1).

The beats in example (1) emphasize important symbolic structures in the spoken construction that are salient to the information structure conceptualization channel: the topical participant in the utterance (Molly **Shannon**), the current discourse topic (**this movie**), and informationally focused elements in the utterance (**unsentimental, brave performance**). A detailed account of informational focus is outside of the scope of this paper, but Langacker (2008: 208) broadly characterizes focus as the components in an expression “which the speaker wishes to foreground as a significant departure from what has already been established in the immediately preceding discourse”. Informational focus is an aspect of information structure in which “new, nonpresupposed information is marked by one or more pitch accents” (Kiss 1998: 246). Halliday (1967:

⁴ There are additional complexities in the gestural channel in this example that we do not discuss. In particular, some of the manual beats are co-expressed with head beats. Head beats are discussed further in example (2).

204) describes information focus as a type of emphasis “whereby the speaker marks out a part of a message block as that which he wishes to be interpreted as informative” (i.e., textually and situationally new information). In (1), we see that the particularly informative elements of objective content at the semantic pole of the spoken composite utterance (i.e., the topic and nonpresupposed information) are emphasized with dependent symbolic structures in gesture and speech (i.e., beats and pitch accents, respectively). Additionally, each of the topical and focused elements in the spoken utterance that are emphasized with beat gestures have salient roles in the stancetaking act that is evoked by the spoken composite utterance.

In looking at the entire utterance, we see that it evokes an evaluative stancetaking act within the information structure channel. Evaluative stance has been defined as an act through which a conceptualizer (the stancetaker) “orients to an object of stance” (the target of the evaluation) and “characterizes it as having some specific quality or value” (Du Bois 2007: 143). Table 1 more clearly distinguishes the components of this particular stancetaking act (adapted from the format used in Du Bois 2007).

Table 1: Whitford Stance Act.

Speaker	Stancetaker	Stance Target	Evaluation
Whitford	Whitford “I” (have to say)	Molly Shannon’s performance in this movie	unsentimental (and) brave

In the utterance immediately preceding the one shown in example (1), Whitford evaluates the authenticity of the film that he is promoting in accurately capturing the experience of losing a parent (the topic of the film). He mentions that he has personally experienced the loss of his own parents and says that he (“I”) “thought this was the most honest (film) including the humor”. In (1), the target of evaluative stance changes from the preceding utterance. The evaluative target moves from the authenticity of the film to Molly Shannon’s performance in the film. This change is indicated by the complement-taking predicate phrase (CTP-phrase) “I have to say”. This collocational CTP-phrase signals that the speaker will present a new evaluation in the complement clause that follows and prepares the addressee for a transition to a new target of evaluation (see Thompson 2002 for an analysis of these CTP-phrases as stance markers). The apex of the first beat aligns phonologically with the complement-taking predicate “say”, which is the most substantive word in the CTP-phrase.

Notice that the speaker-stancetaker in (1), which is elaborated by “I”, does not receive a beat gesture. In this specific example, the stancetaker is not particularly salient. One reason why we argue this to be the case is because the speaker is the expected stancetaker in language use. What is perhaps more important to this particular interaction, however, is that the speaker was already overtly established as stancetaker in the stance act expressed in the preceding utterance. In this case, the speaker as stancetaker is still expected to be active and accessible to the addressee. We will see a different pattern in the second example (2) when different stancetakers are introduced across juxtaposed utterances.

The second beat in (1) aligns with “Shannon”, emphasizing the new target of stance (Molly Shannon) to which the evaluation is oriented. The beat that falls on “Shannon”, rather than Molly, singles out the particular participant at whom the evaluation is directed. It is the word that selects for reference the particular Molly whose performance is being evaluated and distinguishes her from other possible “Mollys” in the world. The next two beats occur on “this movie”, which is the current discourse topic but also part of the object of stance in this evaluation. These two beats, in addition to emphasizing the discourse topic, serve a more localized (i.e., immediate) role in this construction of emphasizing the context or scope in which the evaluative stance of Molly Shannon’s performance is to be interpreted. Whitford is making an evaluation of Molly Shannon’s performance in this particular movie, not an evaluation of her general acting abilities. The fifth and sixth beats align with the evaluative objective content. The attributive expressions “unsentimental” and “brave” assign a favorable evaluation of Molly Shannon’s performance in the (this) movie. The final beat aligns with “performance”, which is also part of the target of the stance act but is temporally separated from the other meanings that are targeted in the evaluation. The full target of evaluation is *Molly Shannon’s performance in this movie*. Whitford’s use of beats in this example emphasizes components in the objective content of speech that serve important roles in the evaluation. The symbolic units in speech that are emphasized with beats function to (1) introduce a new stancetaking act, (2) orient to a new object (or target) of stance, and (3) provide an evaluation.

In (1), the spoken symbolic structures that are highlighted with beats are multifunctional. Not only do these spoken structures serve important roles in a particular stancetaking act that is relevant to the local discourse context; these elements serve key roles at a higher level of information structure, such as the roles of discourse topic and focused participant within that topic. While an utterance includes component elements that have semantic specification in the objective content channel, the significance of the elements to other elements

within a spoken composite construction and the relation that construction has to neighboring constructions, is specified at the information structure channel. The meanings of beats are specified in the information structure channel because their meanings are reliant on the higher-level functions of the component structures in speech. The semantic function of a beat's emphasis is not elaborated by the objective content of the spoken structure with which it is aligned. It is elaborated by the role the spoken structure serves within the composite utterance, across adjacent utterances, and in the broader discourse.

4.2 Analysis 2: beat and points in multimodal assemblies

In this section we examine constructions that incorporate a higher level of complexity in the gestural channel. We analyze multimodal symbolic assemblies that include the simultaneous and integrated expression of beats and pointing gestures.

Our analysis of pointing is also based on CG (Langacker 2016b; Wilcox and Occhino 2016). Pointing falls within the semantic domain of nominal grounding. *Grounding* serves to indicate the epistemic status of a nominal or clausal referent; that is, grounding specifies where a referent “stands in relation to what we currently know and what we are trying to ascertain” (Langacker 2008: 297). For nominals, the goal is to identify a selected referent. Reference is *intersubjective*: nominal grounding permits the speaker and hearer to identify and focus their attention on the same referent from among the multitude of other entities that populate our mental universe. According to Langacker (2016b: 106), “the basis for identification is a **path** – a series of connections – leading from the conceptualizer (the *origin*) to the referent (the *goal* or *target*)”.

Figure 6 depicts a canonical case of pointing. G is the actual ground in the speech event, S and H are the speaker and hearer. The ground includes the area and entities visible to the speaker and hearer, the onstage or objective scene (OS). Pointing (solid arrow) singles out or identifies the one entity on which the speaker wants the hearer to focus her attention (FOC). This act of pointing has a “directive force” (double arrow): “it instructs the hearer to follow its direction, so that both interlocutors end up focusing attention on the same entity, the gesture's intended referent” (Langacker 2016b: 110). In canonical pointing, the directive force of a pointing gesture instructs the hearer to visually follow its direction. In more abstract pointing, the hearer may be instructed to conceptually “follow” the direction so that both interlocutors focus attention on the same conceptual entity.⁵

⁵ It is this use of pointing to direct attention at a conceptual referent that has been extensively grammaticalized in signed languages (Wilcox and Occhino 2016).

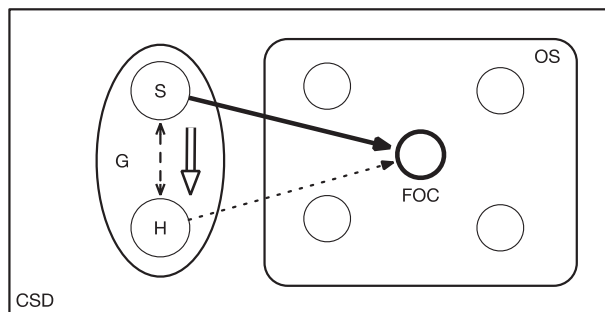


Figure 6: Canonical pointing gesture (after Langacker 2016b).

Langacker’s analysis suggests that pointing consists of two components: a directive force of attention and an entity on which that attention is focused. We thus analyze pointing gestures as complex symbolic structures (i.e., constructions) consisting of two component symbolic structures. The first, a pointing device, serves to direct attention. Its phonological pole is some articulator capable of directing the hearer to follow its direction. Pointing devices vary from culture to culture, the index finger and lax hand being common in western cultures, although any articulator sufficiently mobile enough to be moved or oriented toward some real or imagined entity could serve as a pointing device. The semantic pole of the pointing device symbolic structure is schematic, with a meaning of “direct attention to” or “focus attention on”.

The second component symbolic structure is a Place.⁶ The phonological pole of Place is some location in the ground, the visible physical space surrounding the interlocutors in the current speech event. In a canonical use of pointing it is the location of the referent in the objective situation. In terms of conceptualization, it is the end point of a subjective mental path leading from the conceptualizer to the entity on which the speaker wants the hearer to focus her attention. The semantic pole of Place schematically profiles a *thing*; in an actual speech event it is the intended referent. The two symbolic structures constituting a pointing gesture are thus a gestural construction, as shown in Figure 7. Above the diagram, the double arrow corresponds to Langacker’s directive force which is the semantic function of the pointing device. The bolded circle represents the focus of attention, the entity identified by Place.

As Kendon (2010: 20) has noted, “However pointing is done, all actions regarded as ‘pointing’, seem to have in common a characteristic movement

⁶ Place is capitalized as a reminder that it names the entire symbolic structure.

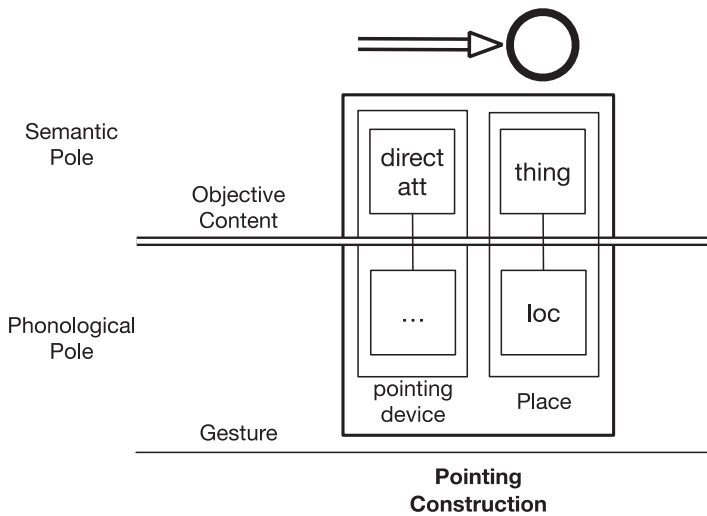


Figure 7: Pointing construction.

pattern in which the body part regarded as doing the pointing is moved in a way that is recognized as having [a] well defined *linear* path.” Conceptually, we suggest that this movement pattern, a dynamic quality described by Eco (1976) as ‘movement toward’, corresponds to the mental path leading from the conceptualizer to the target or intended referent. Phonologically, it is realized as a dependent movement of some more autonomous articulator. The movement, an elaboration of a more autonomous structure, operates on the pointing device’s location, resulting in the linear path ‘movement toward’. Movement also can operate on orientation, producing an orienting rather than linear movement. In either case, the function of the pointing device is to direct attention, conceptually and perhaps visually, at the Place.

The pointing device can integrate with a beat gesture to form a higher lever composite symbolic structure in the gestural channel, a beat-point construction (Figure 8). Together the handshape and movement, either linear or orienting, serve as the phonological carrier for the beat symbolic structure. Because the carrier is a pointing device, it is itself a symbolic structure. The beat, phonologically expressed as manner of movement, modifies the movement of the pointing device. Additionally, the semantic poles of the two symbolic structures are conceptually compatible: the schematic meaning of the pointing device is to direct attention, and the beat has a schematic meaning of emphasis, a kind of focused attention.

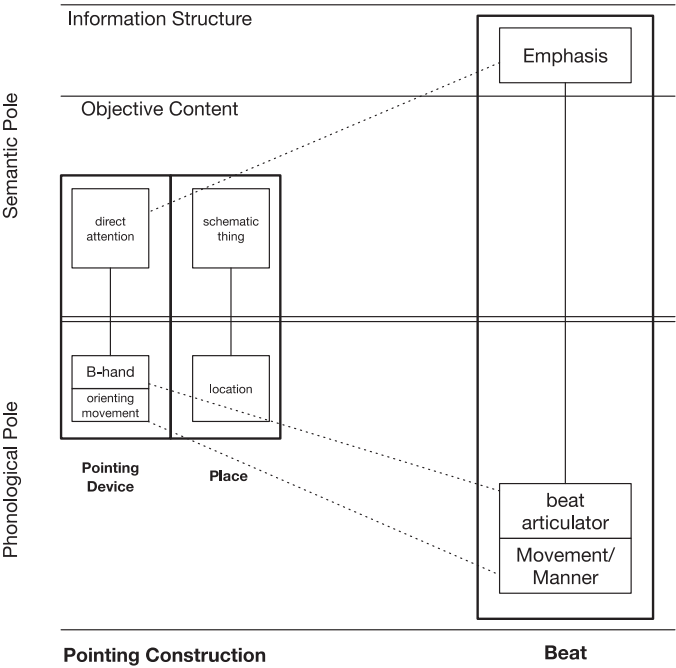


Figure 8: Beat-point construction.

As we will see in our data analysis, pointing constructions can also function to establish reference points. *Reference points* recruit our ability “to invoke the conception of one entity for purposes of establishing mental contact with another” (Langacker 1993: 5). A reference point, labeled R in Figure 9, defines a conceptual region, the dominion (D), which includes all the entities or targets (T) with which the reference point affords mental access. A reference point permits a conceptualizer (C) to establish mental contact with a target. In other words, the conceptualizer’s focus of attention follows a mental path, through the reference point, and makes mental contact with a particular target.

Reference points figure in many constructions, including possessives (Langacker 1993), pronominal anaphora (Van Hoek 1995, 1997), and topics (Langacker 1993), including discourse topics. Reference points are conceptual entities anchoring conceptual structures, their associated dominion (Van Hoek 1997). When used in pointing constructions in signed languages (Wilcox and Occhino 2016), and as we will demonstrate for pointing constructions in gesture, reference points can be realized phonologically as actual points directing attention in the ground, the visible space surrounding interlocutors.

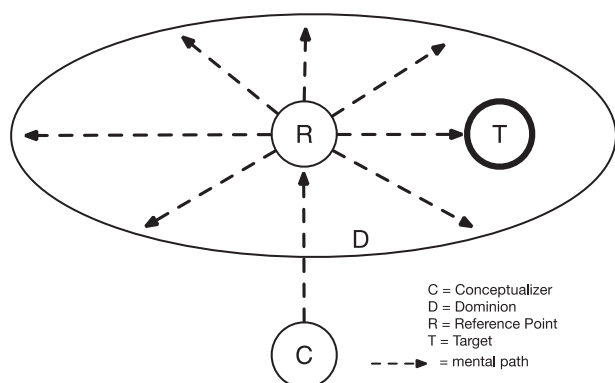


Figure 9: Reference point.

In using a pointing construction as a reference point, the pointing device directs attention to a particular phonological location in the current ground, the phonological pole of a Place structure. Initially, the pointing construction functions in its canonical role to direct attention to a grounded nominal. Once established, any grounded nominal can serve conceptually as a reference point. Subsequent points to this Place structure renew mental contact with the nominal, which now functions as a reference point, instructing the hearer to continue the mental journey to a target in the reference point's (the nominal's) dominion. How does the hearer know which of myriad entities in the reference point's dominion will be the target, the new focus of attention? One strategy is to simultaneously name the target in the speech stream.

Our second example takes place during one of the 2016 presidential primary debates between Hillary Clinton and Bernie Sanders in Flint, Michigan. The candidates are standing on a stage, both behind podiums that are approximately at waist level. From the audience perspective, Clinton is on the left and Sanders is on the right.

During this part of the debate, the audience has been permitted to ask the two candidates questions. The particular segment of the discourse that we have selected to analyze occurs during one of Clinton's turns in responding to the question "How will you encourage companies to keep factories here in the U.S. instead of moving them to other countries?". In Clinton's first turn answering the question, she emphasizes the incentives she plans to give companies who keep their manufacturing plants in the U.S. and describes the penalties she would like to impose on companies that move manufacturing work to other countries. Sanders begins his first turn (following Clinton) by criticizing Clinton for her past voting record, which includes her support of trans-national trade agreements that had provisions in them that Sanders argues were bad for

manufacturing jobs in the United States. Sanders accuses Clinton of “voting for every disastrous trade agreement and voting for corporate America”.

Clinton, in her second turn, responds to Sanders’s criticism by shifting the discourse topic to the auto industry bailout. She criticizes Sanders for not supporting the bailout in early 2009. It is likely that Clinton believes that this topic is especially important to the members of the audience. The people of Flint, Michigan have historically relied heavily on the auto industry for jobs, and the bailout was crucial for the economy of the state of Michigan. Clinton then returns to Sanders’ criticism about her voting record aligning with corporate interests and defends herself by emphasizing her vote on one bill in particular, an auto industry bailout bill negotiated by the Bush administration at the end of his second term in 2009. She acknowledges that there were provisions in the bill with which she did not agree and which she would have written differently, but follows with a play to the audience. Clinton says, “But was the auto bailout money in it [the bill]? The three hundred and fifty billion dollars that was needed to begin the restructuring of the auto industry. Yes, it was.” From this point, we begin our analysis. A transcription of the spoken discourse included in our analysis is shown in (2).⁷

- (2) HILLARY When I talk about Senator Sanders being a one issue candidate,
 CLINTON: I mean,
 very clearly,
 (0.22)
 you have to make hard choices,
 (0.23)
 when you’re in positions of responsibility.
 (0.25)
 The two senators from Michigan,
 (0.26)
 (H) stood on the floor,
 and said we have to get this money released.
 (0.42)
 (H) I went with them,
 and I went with Barack Obama.
 You did not.

We now explore how gestures, particularly beats, points, and beat-point constructions, interact with the spoken language in this example. Appendix 1 shows the conventions used for the labeling of the beat and point gestures. The

7 This broad transcription uses discourse conventions described in (Du Bois et al. 1993).

alignment of beat and pointing gestures with speech are included in the transcription.

Clinton-Sanders Debate Excerpt

Line

1 [So when I][talk ab][out] Senator Sanders
[POINT][POINT HOLD][POINT]

being a *one* *issue* *candidate*,

HEA HEA HEA

2 I *mean*,

HND HEA

3 *very* clearly,

HEA

4 *you* have to make *hard* *choices*,

BOTH

BOTH BOTH

(reduc)

5 when *you're* *in* *positions* of *responsibility*.

HEA HND

BOTH

BOTH

(reduc) (reduc)

(reduc)

(reduc)

(two iterations)

6 [The two *senators*] from Michigan,

BOTH

[POINT]

7 [*stood*] on the *floor*,

BOTH

HND

[POINT]

reduc

8 and [*said*] *we* have to *get* *this* *money* *released*.

HND

HND

HND

HND

HND

HND

[POINT]

reduc

9 [I] went with them,

BOTH

[POINT]

(reduc)

10 and [I] went with Barack Obama.

BOTH

[POINT]

(reduc)

11 [**you**][did not.]
HND
[**POINT**][POINT]
HOLD

In order to account for the symbolic integration of the gestural expressions with the spoken language constructions in this second example, we again need to discuss stancetaking, this time examining stance acts that Clinton performs. We focus first on the initial intonation unit (IU) in the example that is shown in lines 1–5 in the excerpt. In this IU, Clinton establishes a semantic relationship between an evaluative stance that she has established in prior discourse (i.e., Sanders is a one issue candidate) and a new stance act (i.e., people in positions of responsibility are required to make difficult decisions).

Table 2 more clearly distinguishes the components of this stancetaking act. The semantic relationship that is established between the two stances is complex and accessed inferentially. Together the two juxtaposed stances serve two complementary functions: (1) Clinton defending herself against Sanders’s criticism about her voting record and (2) Clinton shifting the criticism to Sanders for simplifying the complexities of bipartisanship.

Table 2: Clinton-Sanders Stance Act.

#	Line(s)	Speaker	Stancetaker	Stance Target	Evaluation
1	1	Clinton	Clinton (“I”)	Bernie Sanders	(is a) one issue candidate
2	4–5	Clinton	Clinton (“I” mean)	people in positions of responsibility (including Clinton)	have to make hard choices

In examining the gestures in the first stance act, we see that Clinton produces a pointing construction on the word “I” in the first clause in line 1 (Figure 10). The pointing construction is phonologically expressed with an open-B handshake, the palm contacting her chest. This pointing construction identifies the semantic pole of the Place with the speaker, as does the spoken first-person pronoun. Clinton’s point is followed by a hold during the expression “talk about” in which her hand remains at her chest in the same location as the phonological pole of the Place that was just established. With this hold, Clinton is continuing to make use of the symbolic Place that she established with the point. The objective



Figure 10: Clinton point to “I”.

content evoked by the co-expressed speech (i.e., the speech during the hold) bears a relationship to the semantic pole of Place. Specifically, Clinton (“I”) semantically assumes the agent role in the process profiled by the verbal expression “talk (about)”. This verbal process (a speaking event) is conceptually dependent on the agent or speaker performing the speaking event (i.e., Hillary Clinton). Clinton also serves as the semantic pole of Place. The hold associated with this Place evokes (through gesture) the relationship between the verbal process and the participant role on which it is dependent (encoded in speech). Following the hold, Clinton performs a second pointing construction in line 1 (Figure 11). This pointing construction is also phonologically expressed with an open-B handshape but here the movement is directed contralaterally to the left side of Clinton’s body (toward Bernie Sanders).⁸ The point slightly precedes the spoken language expression “Senator Sanders” that elaborates Sanders as the semantic pole of Place in this pointing construction. Three head beats follow the points in line 1, aligning with each of the words, *one issue candidate*.

Returning to the discussion of the construction shown in line 1 as a stance act, we see points being assigned to both the stancetaker (Clinton) and the target of stance (Sanders) and beats assigned to the evaluative meanings in the stance act (*one issue candidate*). Each of the key aspects of the stance act are marked

⁸ The different location is somewhat difficult to discern in the static figures. Compare the location of Clinton’s hand relative to her microphone between the two figures.



Figure 11: Clinton point to Sanders.

with attention directing devices in gesture (either points or beats). The beat gestures that align with the evaluation presented in speech are expressed by the head rather than produced manually. Unlike the Whitford example in Section 4.1, there is no beat assigned to the stance target. This example, however, has an important functional difference from the Whitford example; in this case, Clinton is not performing a new stance act. This stance, in which Clinton characterizes Sanders as a “one issue candidate”, has already been established. Clinton is revoking it to serve a new function in relationship to the entirely new stance act that she performs in lines 4–5. The stance act in line 1 is construed as already known by the audience. We can see this expectation reflected in the information structure channel of the speech through the embedding and subordination of this stance act (signaled by the subordinator “when”). It is implied in this use of a temporal adverbial phrase that Clinton has repeatedly expressed this stance about Sanders. Clinton’s recycling of a previously established stance might be the motivation for the low degree of emphasis (at least in the form of beat gestures) that components in this stance act receive.

Separating the first two stancetaking acts in this example, we see a transitional clause (lines 2–3) *I mean, very clearly*. This clause serves a similar function as *I have to say* did in the Whitford example (1). Through the use of the first-person pronoun and the propositional attitude predicate *I mean*, this clause signals that the complement clause that follows will introduce a new stance act. Despite the speakers in the two examples using different semantic types of

complement-taking predicates (an utterance predicate, *say*, and a propositional attitude predicate, *mean*), the clauses are similar in that one of the functions of each is to introduce a new stance act. The predicates in both examples receive a beat gesture (performed with a strong head beat in this second example). This transitional clause *I mean* also participates in the stance act that follows by explicitly identifying the stancetaker as the speaker with the overt use of the first-person pronoun. A manual beat that aligns with the first-person pronoun emphasizes the role Hillary Clinton has as stancetaker in the entirely new stance act that follows. The transitional clause in lines 2–3 is also important because it suggests that the prior stance that Clinton has recalled in line 1 is semantically related to the upcoming stance (lines 4–5). Specifically, the clause, *I mean very clearly* is conceptually dependent on the two stances (the one that precedes it and the one that follows) to have meaning.

The second stance act in this example (lines 4–5) receives beats that emphasize the target (or object) of stance and the evaluation. In this case, the evaluation is introduced before the full target of stance is expressed, and the strongest beats align with words that express the most substantive meanings related to the evaluation (*hard choices*). Beats that are reduced in movement size align with the words that express meanings associated with the target of the stance act (*you/you're, in, positions, responsibility*). The stancetaker and evaluation receive a greater degree of emphasis, in terms of the force of the beat, than the stance target. It is important to note that the stance target is initially construed with a generic use of the second-person pronoun (*you*). This use of generic reference overtly codes the target of stance as a category of people rather than a specific individual. The category includes people in positions of responsibility, a grouping to which Clinton (and Sanders) belong. This particular evaluative stance act has to be interpreted within the context of a previous stance taken by Sanders. Clinton seems to use this stance act to reject alignment with Sanders's prior position in which he suggested that Clinton supports the interests of corporate America. Rather than overtly defending herself in speech (e.g., I had to make hard choices), Clinton construes this evaluation in lines 4–5 as targeting other nonspecific referents who are politicians. By doing this she implies that difficult decisions come with the job of being a politician and that she is not the only politician who has had to compromise on legislation. Perhaps the attenuated beats on the target of stance are related to the fact that Clinton is obscuring the primary target of the stance, herself. Clinton is not overtly coded in speech as the target of stance but can be metonymically understood to be included as part of the target because she is someone in a position of responsibility. The actual target of stance (Clinton) is also retrievable through the relationship this stance act has to the broader discourse, Clinton defending herself against Sanders's earlier criticism.

In the next intonation unit (lines 6–8) we find beats simultaneously integrating not only with meanings expressed in the spoken language but also with pointing constructions to form symbolic assemblies in gesture. In line 6, Clinton directs a point on the words *the two Senators* to a location in the audience, fully extending her arm toward that location, which is the phonological pole of the pointing construction's Place structure. It is perhaps also the physical location of the Senators, who may well have been present for the debate and seated in the audience. The schematic semantic pole of this Place is elaborated by the spoken phrase *the two Senators* (Figure 12a). As the pointing device reaches the phonological location of the Place, Clinton simultaneously performs a strong beat gesture with both her head and hand. The beat aligns with the *senators*, who function discursively as both topical participants as well as stancetakers in the stance act that is presented in a clause that follows (line 8). Clinton was the stancetaker in the prior stance act. This strong emphasis on *senators* may, at least in part, be due to the fact that they are being established as new stancetakers (in addition to the important relationship the senators have to the audience). Subsequent points are directed to the same location, with beats also occurring on the words *stood*, and *said* (Figure 12(b) and 12(c)). Each beat-point is attenuated phonologically, both in the force of their downward movements and in the degree to which the arm is extended. The last beat, on the word *said*, is the least extended of the three; this beat is also attenuated, produced only with the hand.

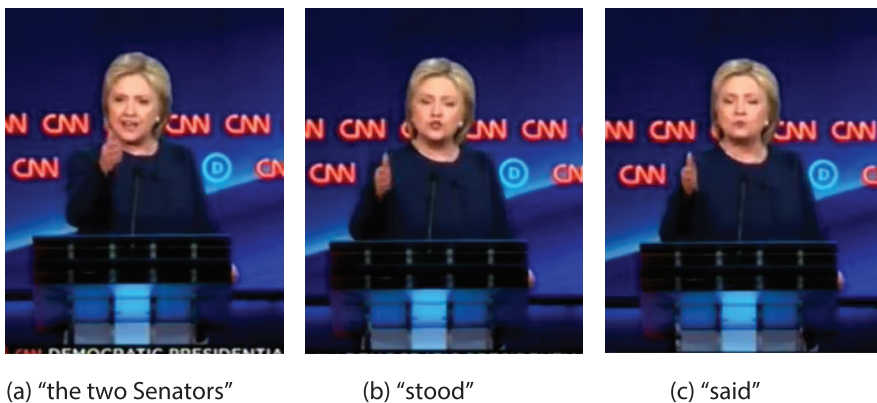


Figure 12: Clinton "the two senators stood on the floor and said..."

There is a notable semantic difference between the first Place structure that is emphasized and the subsequently emphasized Place structures. In the first, the schematic semantic pole of the Place structure, a thing, is elaborated by the

spoken words *the two Senators*. In the subsequent two pointing constructions, the semantic pole of Place is not elaborated by the simultaneous spoken words: the semantic poles of these two Places are not ‘stood’ and ‘said’. Rather, by directing the pointing device to the same Place already used in the pointing construction associated with *the two Senators*, these points refer anaphorically to “the two Senators” as a reference point, and its dominion provides “a context with respect to which an expression is interpreted (or into which its content is integrated)” (Langacker 1993: 24). The spoken words in this case evoke targets in the dominion of the reference point: they identify salient actions that the senators perform as members of Congress: *standing* (on the floor) and *saying* are actions they took to garner support for the passage of legislation to release the bailout funds. These actors and their actions are also important to Clinton’s defense of her position on the bailout (voting in favor of it) and her continued response to Sander’s criticism. The senators, with whom she expects her audience to align, publicly held the same position that she did. In the composite gestural construction, the point and beat work together to establish a reference point and to emphasize that reference point (and its dominion) as being particularly salient in the discourse, for the reasons we have discussed.

At this point, Clinton changes the direction of movement, and in our analysis, although she continues to produce a series of rapid beats with the B-handshape, the hand no longer serves as a pointing device in a pointing construction. The direction of her arm moves from central or neutral gesture space along a linear path towards Clinton’s right. A quick, reduced beat occurs on *we*, with stronger beats occurring on *get*, *this*, and *money*. The beats are short, downward movements along the rightward path. The path movement ends with Clinton’s arm extended at a final location at her right on the word *released* (Figure 13). This path movement can be interpreted as metaphorically mapping the concept of time onto space using a transversal temporal timeline (i.e., one that runs horizontally), as has been observed in co-speech gesture in some languages (Casasanto and Jasmin 2012; Cooperrider and Núñez 2009). In particular, English speakers typically gesture from left to right to show development through time when using a transversal timeline. In this example, the event being described in the complement clause of the speech event, “get this money released”, emphasizes the protracted process leading up to the release of the money. This movement from left to right during the expression of the beats with the complement clause may be interpreted as evoking that development through time. The placement of beats on *we*, *get*, *this*, *money* and *released* along with the path motion seem to emphasize the steps in the process of getting the money released by aligning with the most salient meanings in the event structure (the agent, the patient and the event). The composite multimodal construction emphasizes the process and steps directed at



Figure 13: “We have to get this money released”.

achieving the end goal, which is the release of the bailout funds. At the same time, the complement clause elaborates on the stancetaking act by introducing the target and the evaluation of the stance act taken by the two senators (Table 3). Again, the beats are multifunctional in the emphatic role they serve in the multi-modal composite, with one dimension of their meaning corresponding (now familiarly) to the emphasis of roles associated with stancetaking.

Table 3: Clinton-Senators Stance Act.

#	Line(s)	Speaker	Stancetaker	Stance Act Target	Evaluation
3	6–8	Clinton	the two senators from Michigan (in 2009)	the money for the auto industry bailout	should be released

In lines 9–11, Clinton again produces pointing constructions simultaneously with beats expressed on both the hands and head. The first occurs in the phrase “I went with them” (Figure 14). One might think that the pointing construction co-occurring with *I* would be directed towards the speaker, as we saw earlier in this example (line 1). Here, however, the B-handshape that serves as the pointing device and manual beat carrier on the word *I* (line 9) is directed toward the location associated with the Place previously established for the senators and their stance. Clinton continues to use this location in her next statement, “and *I* went with Barack Obama,” again placing a point and a beat on *I*. While the arm



Figure 14: Clinton “I went with them”.

extension associated with the subsequent point is reduced or attenuated, the manner of movement associated with the beat is strong, being co-expressed manually and with the head. Conceptually, by saying, “I went with them and I went with Barack Obama”, Clinton has aligned herself with the stance of the two senators (“them”) and their efforts to get the money for the auto industry bailout released, which she established in the previous intonation unit. President Obama’s stance (which Clinton established as being in support of the bailout several minutes prior in the discourse) is also shown to be included in the dominion of the reference point that was previously established in that location. She accomplishes alignment phonologically in gesture by using the same location, a physical location in the audience, across all these Place structures. Through this use of the phonological location associated with the reference point in gesture, Clinton also aligns herself conceptually with the voting members of the audience, many of whom undoubtedly were workers in the auto industry and who supported their state senators and Obama’s efforts to get the

bailout approved. She further emphasizes this alignment through the co-expression of beats with the pointing construction.

This conceptual and phonological alignment is then used to create a vivid contrast in line 11. During this IU, Clinton changes arms and produces pointing gestures with her left arm. First, she directs the B-handshape pointing device on her left hand toward a Place on her left while saying, “**You** did not” (Figure 15). The phonological pole of this Place coincides with the physical location of Sanders, who is standing on her left. The semantic pole of Place, which is elaborated by the second-person pronoun “you”, is Bernie Sanders. Contrasting with the Place used for the two senators, Obama, and her alignment with them, Sanders is conceptually placed in opposition to the bailout and its supporters; by switching arms and pointing to Sanders with a strong beat manner of motion, she strengthens the contrast by using his location on stage as the phonological location of the gestural Place structure produced with *You*. Through this complex use of speech and gesture, Sanders is portrayed as being politically, conceptually, and physically (phonologically) up on stage, aloof as it were, and non-aligned with the auto industry bailout, the politicians who supported it, and most importantly, the audience. On the other hand, Clinton has portrayed herself as politically, conceptually, and physically (phonologically) aligned with the voters, their senators, and the president who supported the restructuring and rescue of their employer, the auto industry.



Figure 15: Clinton “You did not.”.

The complex beat-point constructions in lines 9–11 are used to establish and emphasize relationships between different stances, and by extension, different stancetakers. Beat-point constructions align with spoken symbolic structures that evoke the speaker and hearer at the semantic pole (*I* and *you*). In order to understand the significance of the repeated use of this schematic gestural construction (i.e., the beat-point construction) in these specific multimodal expressions, one has to consider the roles of the speaker and hearer at both the local level (within the specific utterance and neighboring utterances) as well as their role in the broader discourse context. If we look at the general discourse genre (i.e., Presidential debate) and context of the speaking event (in front of an audience from Michigan), we can recognize that the goal of this discourse is for Clinton to convince the audience of eligible voters and other eligible voters watching remotely that she is on their side and that, ultimately, they should vote for her rather than Bernie Sanders. Within this particular section of the discourse Clinton is making an appeal to the audience, hoping that they will recognize that she has stood with them in the past when Sanders has not. The overall aim of a Presidential debate is for the candidates to gain new supporters. If one considers the relationship between that aim and the local linguistic context (lines 9–11) in which Clinton shows her alignment with the voting audience and Sanders' lack of alignment with the voting audience, the contrast in alignment between the two candidates is especially salient to the discourse. This saliency is overtly encoded in the gestural channel in the form of the magnified attention directing beat-point construction. One could analyze an emphasized point to Bernie Sanders as explicitly referring to Bernie Sanders, but it is the role that Sanders (evoked in speech by “you”) serves in the clausal context (topic of clause), the interactional setting (political opponent), and his relationship to the stances evoked in previous constructions (not aligned) that all interact with the use of this beat-point construction.

5 Discussion: directing intersubjective attention

Writing on the significance of intonation, Bolinger (1986: 74) pointed out that:

Logical people like to view language as primarily the business of exchanging information. This view is reinforced by the importance we attach to writing: most of what we read is written to inform, either the mind or the imagination. But speech is *différent*. It informs sometimes (as often inadvertently as by intent), but much of the time its aim is to cajole, persuade, entreat, excuse, cow, deceive, or merely to maintain contact – to let the hearer know that ‘channels are open’. Furthermore, even when we inform we are not above slipping in an extra message *sub rosa*: ‘the information I am giving you is important’. (Bolinger 1986: 74)

Speaking, gesturing, and signing provide language users with symbolic resources with which they strive to share a conceptual world and jointly focus attention on entities and occurrences within that shared world. If this can be accomplished, speakers and interlocutors are said to have achieved *intersubjective alignment* – “momentary alignment in the interlocutors’ scope of awareness and focus of attention” (Langacker 2017: 2). Although an auspicious accomplishment, joint attention alone is not enough. Intersubjective alignment requires more: the sub rosa message “the information I am giving you is important”. Not all of the information is equally important in terms of the contribution it makes to a speaker’s communicative goals. A variety of symbolic resources, such as word choice, intonation, beats, and points can be recruited by the speaker to direct an addressee’s attention to those components of an expression that are especially important.

As Langacker (2001) has pointed out, another dimension of linguistic organization serving to direct attention at the level of discourse presents information in “windows of attention” marked by intonation units (Chafe 1994). Speakers use these *attentional frames* to “regulate the amount of effortful material presented in a single window of attention” (Langacker 2001: 154–155). This in turn facilitates the management of attention. Attentional frames are also symbolic structures: the phonological pole consists of an intonational grouping which serves particular functions associated with the information structure channel at the semantic pole.

Langacker (2017) writes that grammar refâlects the interplay of *descriptive* and *discursive* factors, each of which involves focusing or directing attention. Descriptive focusing involves profiling or conceptual reference. For example, a noun such as *Senators*, when occurring in the nominal clause *The two senators*, identifies and profiles a conceptual referent in the objective content channel. Pointing serves much the same referential function, identifying and profiling some entity towards which the speaker directs the attention of the hearer. Discursive focusing is a higher-level analog of focusing. Discursive focus is often expressed through the information structure channel. In speech, discursive focus may be expressed as pitch accent or attentional frames. Beats are similar to attentional frames in that their meanings are associated with the information structure channel.⁹ Unlike attentional frames, however, which typically incorporate larger sequences of structural elements within an intonation unit, beats

⁹ Although specified in different vocalization channels, beats and attentional frames are similar at the phonological pole as well: there is a well-known connection between intonation (the phonological pole of attentional frames) and gesture (the phonological pole of beats) (Bolinger 1983, 1986).

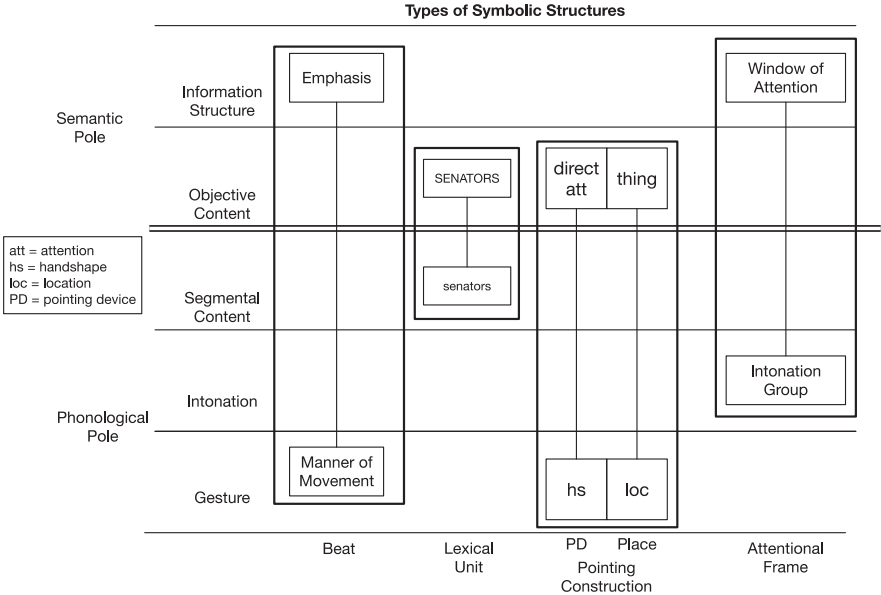


Figure 16: Spoken and gestural symbolic structures.

operate on a narrower scope. In multimodal speech-gesture constructions, beats align with pitch-accented syllables; when co-expressed with pointing constructions in the gestural channel, they occur on the pointing device. Figure 16 illustrates these symbolic structures and how they are manifest in different conceptualization and vocalization channels.¹⁰

Recognizing that symbolic structures can be expressed through different vocalization channels and can have meanings that are most salient to a particular conceptualization channel is useful because it allows us to better account for how and why symbolic structures are integrated in multimodal constructions. In constructions that incorporate simple beats and speech (as discussed in 4.1), the endpoint of the movement that elaborates a beat in the gestural channel aligns with pitch accents in the intonation channel. At the semantic pole, beats and pitch accents are both markers of emphasis (conventionally called “prominence” for pitch accents), which is a broad category of meaning that is associated with the information structure conceptualization channel. In specific usage events, pitch accents are functionally associated with focus

10 For simplicity, the phonological pole of a pointing device is shown only as a handshape. Our diagram is an elaboration of Langacker 2001, Figure 12.

(Pierrehumbert 1980; Selkirk 1986). Both beats and pitch accents are semantically dependent on speech because their emphatic meanings in the information structure channel require material from the objective content channel (i.e., the thing or relationship being emphasized and focused). Pitch accents are also phonologically dependent on segmental content in speech as their expressive carrier. We saw that beats sometimes provide additional emphasis to informationally focused meanings to which pitch accents also direct attention. Multimodal constructions that incorporate beats were additionally found to have a relationship to other meanings elaborated through information structure. We saw, for example, that beats emphasize segmental content that, in addition to having objective content meaning, serves higher-level functions as the major components of stancetaking acts.

In more complex multimodal constructions in which beats and points are co-expressed, the beat apex not only aligns with the pitch accents in the intonation channel of speech but with the pointing device in the gestural channel. Even though beats and pointing gestures are expressed through the same vocalization channel, their phonological co-expression is possible because the pointing device does not require a particular manner of movement. The manner of movement aspect is thus available for the expression of beats. At the semantic pole, the co-expression of beats and points is also felicitous: they are both used to direct intersubjective attention. When beats and pointing constructions are co-expressed as a complex gestural construction with speech, we see these attention directing devices integrate to serve unique functions. In the Clinton-Sanders debate, for example, they were used in coordination to provide extra emphasis when establishing reference points elaborated through the spoken channel.

While multimodal constructions that incorporate pointing can establish and direct attention to reference points (Place structures) without the help of beats, the integration of beats and points emphasizes that certain stancetakers and their stances are particularly important to the speaker's message. In particular, we saw that Clinton's use of the beat-point constructions heightened the amount of attention directed to various stancetakers and their relationship (i.e., aligned or not aligned) to the stance with which the audience was expected to assume. Together, beats and points provide a higher degree of directive force than when points are expressed without beats. Patterns in the data suggest this added directive force, which instructs the hearer how to "apprehend [a message] in accordance with established convention" (Langacker 2008: 460), may be motivated in part by the performance of an important functional component of stancetaking: positioning and alignment of stancetakers across stance acts. While beat-point constructions can increase the directive force they lend to the

symbolic structures with which they align, what they don't do is directly reveal why those structures are important. It is up to the addressee to interpret the significance from the larger discourse context as it unfolds in time.

6 Conclusions

Directing attention is a ubiquitous conceptual function in discourse. Attention directing strategies are available in both spoken language and gesture. We have examined two strategies that are expressed in the gestural channel, beats and points. Points serve this semantic function through the directive force of the pointing device, which instructs the hearer to direct attention at the Place, specifically at the entity that elaborates the semantic pole of Place. Beats direct attention through their semantic function of emphasis in the information structure conceptualization channel.

The semantic and phonological poles of symbolic structures exhibit internal complexity, described in CG as the conceptualization and vocalization channels. What has not been recognized to date is that the gesture channel is internally complex. Manual gestures are composed of handshapes in certain locations and orientations, which move with a particular manner of movement. These articulatory aspects of gestures exhibit autonomy/dependence relations. Further, we have shown that these aspects of the gestural channel may independently become associated with semantic content. We have demonstrated, for example, that beats are symbolic structures, their semantic pole manifest in the information structure channel and their phonological pole expressed as manner of movement.

Beats are multiply dependent symbolic structures. Being expressed as manner of movement, beats are highly dependent structures in the gesture channel, making schematic reference to movement, which itself is dependent because it makes schematic reference to an autonomous articulator (hand, head, etc.). Beats are also dependent on the phonological pole of symbolic structures expressed in spoken language. Beats align with and are dependent on pitch accents specified in the intonation channel, which in turn are dependent on segmental content.

Beats are also highly dependent semantic structures. Beats make reference to some autonomous content, the information that is emphasized or highlighted. This autonomous content is typically elaborated by the objective content channel of the semantic pole of the accompanying speech. As we saw in our data analysis, because the accompanying speech also consists of complex constructions integrating objective content with information

structure, and because beats are semantically dependent on speech, beats exhibit high degrees of semantic complexity as well. It is this highly dependent and complex symbolic character of beats that make them particularly well-suited to combine with other gestures, such as points. As we noted, while other researchers (McNeill et al. 2015) have observed that beats are often “overlaid” on other gesture types, our analysis reveals how this semantic and phonological integration takes place. Perhaps the multiply dependent phonological and semantic properties of beats are also what have led to their characterization as “formless hand movements that convey no semantic information” (Özçalışkan and Goldin-Meadow 2009: 3).

As analysts, we often choose to focus on specific dimensions of meaning because of the need to narrow the scope of our analysis. For instance, we can zoom in and look at the objective content meaning of a noun within a noun phrase, or zoom out and look at the participant role the noun phrase serves within the clause, such as agent or patient. We can zoom out further to examine the role of the noun phrase across neighboring clauses, or even further to look at the function the noun phrase serves at a higher level of discourse. While we often focus only on one or two levels of meaning in a particular study, it is important to recognize that in interaction a noun/noun phrase (or any type of expression) evokes all of these dimensions of meanings at once. In language use, symbolic structures can express meanings that are associated with objective content, but those structures are invariably structured in relationship to one another, serving particular purposes associated with information structure and speech management. In focusing on the information structure channel, we have found that points and beats in composite multimodal constructions can be used to emphatically establish references that are associated with stancetakers and their positioning. We also saw that gestural reference points that have previously been associated with a particular stancetaker and stance act can be repurposed with a beat-point construction to show the speaker’s alignment with that stancetaker and their position.

It is important to recognize that this study was microanalytic, and the patterns observed in the data cannot be generalized to suggest that beat constructions or beat-point constructions are used only or typically for these functions. However, our findings provide a useful point of departure for future investigations, which would test the patterns we observed across a greater number of tokens and speakers and across other genres. While we have only examined beat gestures and beat-point constructions, we believe our analysis could be applied to other gestural constructions integrating, for example, beats and cyclic gestures (Ladewig 2012), or points and palm-up-open-hand gestures (Müller 2004).

Some questions that arise from this study are (1) are beats typically associated with the emphasis of elements of stancetaking, and (2) do beat-point constructions serve functions other than those related to emphatically marking stance alignment and positioning? Additionally, there was variability in the degree of emphasis provided by beats, such as the difference in emphasis provided by a head beat alone versus a head beat that is co-expressed with a manual beats. While we speculated on motivations for the cases observed in the data, future research should examine whether there are significant functional patterns to the use of beat gestures of differing articulators and intensities.

An important implication that arises from this study is an understanding that manual gestures, which have historically been regarded as holistic and non-componential, can be symbolically complex gestural constructions, as we saw with beat-point constructions. We urge multimodal language researchers not to define a complex gestural construction by the role that a single symbolic structure plays within that construction. We include as constructions those that integrate component spoken symbolic structures (spoken constructions); those that integrate component gestural symbolic structures (gestural constructions); and constructions composed of component symbolic structures from speech *and* gesture. The latter are multimodal constructions.

We hope to have demonstrated in this paper that gesture can be described using the same theoretical and analytical concepts that Cognitive Grammar applies to the study of language, and that by taking this approach reveals much about how speech and gesture are integrated at all levels of discourse. The overall result, we believe, will lead to conceptual unification and theoretical clarification in our understanding of the relation between language and gesture.

Appendix 1: Methods

Each video segment included in the analysis was exhaustively examined for manual and non-manual beat gestures using ELAN. Each potential manual beat token was examined using a frame-by-frame analysis (first described by Seyfeddinipur 2006: 104–106). For each token, the first frame in which the hand was blurred on the downstroke was marked as the beginning of the stroke phase and the last frame in which the hand was blurred on the downstroke was marked as the end of the stroke phase. A stroke is the primary movement phase that characterizes a dynamic gesture. To be counted as a manual beat, hand blurring had to be present on the downstroke. The beat apex was identified as the furthest point on the downstroke and was

characterized by a lack of blurring of the hand. Non-manual beats were identified as those in which the head moved abruptly forward stopped before retraction. In order to be coded as a non-manual beat, a further requirement was that the stroke that included the beat be expressed on a prosodically stressed content word.

Separate tiers were created in ELAN to code handshapes and the spoken language that corresponded to the stroke. After the initial coding process by the first author, both authors examined each of the tokens to ensure that they agreed on each token’s coding status as a beat. As non-manual beats do not reliably show blurring of the articulator during the stroke phase, this agreement between researchers was a particularly important step. This same frame-by-frame analysis was followed for the identification of pointing gestures.

In the transcriptions, the bolded words align with beats. The italicized syllable in a bolded word aligns with the apex of the beat.

Appendix 2: Gesture transcription conventions

Beats

manual beats = HND	words that align with manual beats are bolded
head beats = HEA	words that align with head beats are underlined
manual and head beats co-expressed = BOTH	words that align with co-expressed manual and head beats are both bolded and underlined
apex of beats	the syllable(s) with which a beat apex aligns is/are italicized

Points

POINT	includes the period of movement toward a location (that corresponds to the PLACE at the semantic pole) and the endpoint of the movement (when the location/PLACE is reached); boundaries are denoted by []
POINT HOLD	static phase following a point in which the hand remains in the phonological location that corresponds to the semantic PLACE

Other Conventions

(reduc)	the gesture is noticeably reduced in size of movement
---------	---

References

- Alexanderson, S., D. House & J. Beskow. 2013. Extracting and analysing co-speech head gestures from motion-capture data. In Robert Eklund (ed.), *Proceedings of Fonetik 2013, the XXVth Swedish Phonetics Conference (Studies in Language and Culture 21)*, Linköping University Electronic Press 1–4.
- Alibali, M. W., D. C. Heath & H. J. Myers. 2001. Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language* 44. 169–188.
- Battison, R. 1978. *Lexical borrowing in American Sign Language*. Silver Spring, MD: Linkstok Press.
- Biau, E. & S. Soto-Faraco. 2013. Beat gestures modulate auditory integration in speech perception. *Brain and Language* 124. 143–152. 10.1016/j.bandl.2012.10.008.
- Biau, E. & S. Soto-Faraco. 2015. Synchronization by the hand: The sight of gestures modulates low-frequency activity in brain responses to continuous speech. *Frontiers in Human Neuroscience* 9. 527. 10.3389/fnhum.2015.00527.
- Biau, E., M. Torralba, L. Fuentemilla, R. De Diego Balaguer & S. Soto-Faraco. 2015. Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. *Cortex* 68. 76–85.
- Bolinger, D. 1983. Intonation and gesture. *American Speech* 58. 156–174.
- Bolinger, D. 1986. *Intonation and its parts: Melody in spoken English*. Stanford, CA: Stanford University Press.
- Casasanto, D. & K. Jasmin. 2012. The hands of time: Temporal gestures in English speakers. *Cognitive Linguistics* 23(4). 643–674.
- Chafe, W. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Clark, H. H. 2003. Pointing and placing. In S. Kita (ed.), *Pointing: Where language, culture, and cognition meet*, 243–268. Mahwah, NJ: Psychology Press.
- Cooperrider, K. & R. Núñez. 2009. Across time, across the body: Transversal temporal gestures. *Gesture* 9(2). 181–206.
- Du Bois, J. W. 2007. The stance triangle. In Robert Englebretson (ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction*, 139–182. Amsterdam: John Benjamins.
- Du Bois, J. W., S. Schuetze-Coburn, S. Cumming & D. Paolino. 1993. Outline of discourse transcription. *Talking Data: Transcription and Coding in Discourse Research*, 45–89.
- Eco, U. 1976. *A theory of semiotics*. vol. 217. Bloomington, IN: Indiana University Press.
- Goodwin, C. 2003. Pointing as situated practice. In S. Kita (ed.), *Pointing: Where language, culture and cognition meet*, 217–241. Mahwah, NJ: Psychology Press.
- Halliday, M. 1967. Notes on transitivity and theme in English II. *Journal of Linguistics* 3. 199–244.
- Holle, H., C. Obermeier, M. Schmidt-Kassow, A. D. Friederici, J. Ward & T. C. Gunter. 2012. Gesture facilitates the syntactic analysis of speech. *Frontiers in Psychology* 3. 74.
- Kendon, A. 2010. Pointing and the problem of “gesture”: Some reflections. *Rivista Di Psicolinguistica Applicata* 10. 19–30.
- Kiss, K. É. 1998. Identificational focus versus information focus. *Language* 74(2). 245–273.

- Kita, S. 2003. Pointing: A foundational building block of human communication. In S. Kita (ed.), *Pointing: Where language, culture, and cognition meet*, 1–8. Mahwah, NJ: Psychology Press.
- Kok, K. I. & A. Cienki. 2015. Cognitive grammar and gesture: Points of convergence, advances and challenges. *Cognitive Linguistics* 27(1). 67–100.
- Krahmer, E. & M. Swerts. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language* 57. 396–414.
- Ladewig, S. H. 2012. Putting the cyclic gesture on a cognitive basis. *CogniTextes. Revue de l'Association Française de Linguistique Cognitive* 6. 1–22. 10.4000/cognitextes.406.
- Langacker, R. W. 1987. *Foundations of cognitive grammar: Volume I, Theoretical foundations*. Stanford: Stanford University Press.
- Langacker, R. W. 1993. Reference-point constructions. *Cognitive Linguistics* 4. 1–38.
- Langacker, R. W. 2001. Discourse in cognitive grammar. *Cognitive Linguistics* 12. 143–188.
- Langacker, R. W. 2008. *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.
- Langacker, R. W. 2016a. Baseline and elaboration. *Cognitive Linguistics* 27. 405–439.
- Langacker, R. W. 2016b. *Nominal structure in cognitive grammar*. Lubin, Poland: Marie-Curie Skłodowska University Press.
- Langacker, R. W. 2017. Evidentiality in cognitive grammar. In J. I. Marín-Arrese, G. Haßler & M. Carretero (eds.), *Evidentiality revisited*, 13–55. Amsterdam: John Benjamins.
- Leonard, T. & F. Cummins. 2011. The temporal relation between beat gestures and speech. *Language and Cognitive Processes* 26. 1457–1471.
- Loehr, D. P. 2004. *Gesture and intonation*. Washington, DC: Georgetown University dissertation.
- McNeill, D. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, D., E. T. Levy & S. D. Duncan. 2015. Gesture in discourse. In Deborah Tannen, Heidi E. Hamilton & Deborah Schiffrin (eds.), *Handbook of discourse analysis*, 262–290. Oxford: Blackwell.
- Müller, C. 2004. Forms and uses of the palm up open hand: A case of a gesture family. In R. Posner & C. Müller (eds.), *The semantics and pragmatics of everyday gestures*, 234–256. Berlin: Weidler.
- Özçalışkan, S. & S. Goldin-Meadow. 2009. When gesture-speech combinations do and do not index linguistic change. *Language and Cognitive Processes* 24. 190–217.
- Pierrehumbert, J. B. 1980. *The phonology and phonetics of English intonation*. Cambridge, MA: MIT Press.
- Pisoni, D. B. 1997. Some thoughts on “normalization” in speech perception. In K. Johnson & J. W. Mullennix (eds.), *Talker variability in speech processing*, 9–32. San Diego, CA: Academic Press.
- Selkirk, E. O. 1986. *Phonology and syntax: The relationship between sound and structure*. Cambridge, MA: MIT Press.
- Seyfeddinipur, M. 2006. *Disfluency: Interrupting speech and gesture*. Nijmegen: Radboud University Nijmegen dissertation.
- Stickles, E. 2016. *The interaction of syntax and metaphor in gesture: A corpus-experimental approach*. Berkeley, CA: University of California dissertation.
- Stokoe, W. C. 1960. *Sign language structure (Studies in Linguistics, Occasional Papers. vol. 8)*. Buffalo, New York: Department of Anthropology and Linguistics, University of Buffalo.

- Theune, M. & C. J. Brandhorst. 2010. To beat or not to beat: Beat gestures in direction giving. In S. Kopp & I. Wachsmuth (eds.), *Gesture in embodied communication and human-computer interaction*, 195–206. Berlin & Heidelberg: Springer Verlag.
- Thompson, S. A. 2002. “Object complements” and conversation towards a realistic account. *Studies in Language. International Journal Sponsored by the Foundation “Foundations of Language”* 26(1). 125–163.
- Tuggy, D. 2007. Schematicity. In D. Geeraerts & H. Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 82–116. Oxford: Oxford University Press.
- Van Hoek, K. 1995. Conceptual reference points: A cognitive grammar account of pronominal anaphora constraints. *Language* 71. 310–340.
- Van Hoek, K. 1997. *Anaphora and conceptual structure*. Chicago: University of Chicago Press.
- Wilcox, S. & C. Occhino. 2016. Constructing signs: Place as a symbolic structure in signed languages. *Cognitive Linguistics* 27. 371–404.