Chapter 2. A theoretical motivation for incorporating gesture into constructions

2.1 Introduction

The need for research on multimodal constructions has been noted in quite a few studies, such as Cienki (2015, 2016, 2017b), Kok & Cienki (2016), Schoonjans (2014 a, b; 2017), Steen & Turner (2013), Turner (2017), Ziem (2017), Zima (2014), and Zima and Bergs (2017). The most extensive and comprehensive discussion can be found in Kok and Cienki (2016), which nicely indicates how the notion of gesture is compatible in nature with the concept of construction. This compatibility could lay a foundation for the proposal of multimodal constructions. Yet, Kok and Cienki (2016) has mainly focused on the aspects of symbolicity, schematicity, and various dimensions of construal in constructions, by taking Cognitive Construction Grammar (Langacker 1987, 1990, 1991, 2008a) as a starting point. It follows that some other basic tenets in construction grammars, such as the holistic nature of constructions and inheritance relations among constructions, have largely been untouched. Therefore, the present chapter aims to continue the discussion on how properties of constructions and gestures are consistent with each other, by considering more basic tenets shared in construction grammars. This could furthermore show how incorporation of gesture into constructions is theoretically necessary and possible, thereby providing the theoretical motivation for multimodal constructions, which lays a theoretical foundation for the present thesis.

Although construction grammars comprise various approaches, such as Berkeley Construction Grammar (Fillmore et al. 1988; Kay & Fillmore 1999), Lakoffian and Goldberg's Construction Grammar (Goldberg 1995, 2006), Cognitive Grammar (Langacker 1987, 1990, 1991, 2008a), Radical Construction Grammar (Croft 2001), and Embodied Construction Grammar (Bergen & Chang 2005), etc., they share the following basic tenets: a) constructions are building blocks of language; b) lexicon and syntax form a continuum, lacking a clear boundary between them; c) constructions are organized in a network via various inheritance links. Furthermore, anchored in cognitive linguistics, many construction grammars take a usage-based approach (Bybee 2006; Croft 2001;Goldberg 2003, 2006; Langacker 1987, 2008a) and view meaning as embodied (Bergen & Chang 2005) and as a form of conceptualization (Goldberg 1995, 2006; Langacker 2008a) (see also Hoffmann and Trousdale 2013 for an overview). All these tenets in relation to gesture will be discussed below.

This chapter is organized as follows. Section 2.2 starts off by introducing a usage-based model, since this offers an initial call for incorporation of gesture. Section 2.3 then discusses the last two tenets above – the embodied nature of meaning and meaning as equal to conceptualization together with the tenet of constructions as building blocks of a language, since they all concern the properties of an individual construction. Section 2.4 is concerned with differences among constructions (of different levels of complexity and schematicity), while Section 2.5 considers the relations among constructions. At the end, I will provide a brief summary as well as possible ways to study the multimodality of constructions, following these basic tenets in construction grammars.

2.2 A usage-based approach to constructions

A usage-based approach assumes that grammar originates from language use (Barlow & Kemmer 2000; Bybee 2006; Bybee & Hopper 2001; Goldberg 2003, 2006; Langacker 2001, 2008a). Without distinguishing between competence and performance, it holds that any linguistic aspect in language use, either abstract or concrete, can become conventionalized or entrenched, and thus can become part of grammar. Furthermore, frequency is an important concept in this approach. Langacker (2001: 146) holds that any "recurrent" aspect of a usage event could emerge as a linguistic unit. Similarly, Goldberg argues that any linguistic patterns can gain the status of construction "as long as they occur with sufficient frequency" (2006: 5). In all, usage-based construction grammars require that a linguistic study should consider language use, and in particular, frequently used patterns.

What then constitutes the use of language? As Carter and McCarthy (2015:11) point out, "spoken language is visual and gestural and not simply a textual phenomenon". Gesture is actually ubiquitous and essentially an integral part of language use, the evidence for which includes the fact that a) people gesture across cultures (Goldin-Meadow 1999); b) people gesture when they talk at various ages in the lifespan, either as children or adults (Goldin-Meadow & Butcher 2003); c) people gesture in various situations, whether to a hearing or non-hearing listener (Iverson & Goldin-Meadow 1998), and whether in a face-to-face situation or over the phone (Bavelas et al. 2008); d) a large variety of discourse and pragmatic functions in spoken language can be expressed by gesture, such as signaling turn taking (Streeck 2009a; Streeck & Hartge 1992), co-constructing a turn (Yasui 2013), or topicality and topic shift (Quek et al. 2002), foregrounding and backgrounding (Enfield 2004; Enfield et al. 2007; Engberg-Pedersen 2011), referent tracking (Debreslioska et al. 2013), coordination and subordination (McNeill & Levy 1993), and many others.

Besides the use of gesture in a large variety of contexts, a number of recent

studies have shown that gestures frequently co-occur with specific verbal constructions. Initial examples include studies on motion constructions (Zima 2014), discourse markers (Schoonjans 2014), Tense-Aspect-Modality markers (Hinnell 2014), and the existential construction-*There be* (Mittelberg 2017), etc. Furthermore, Sekine and Kita (2015) observed that, when asked to retell stories about motion events demonstrated in video clips, participants performed on average approximately one representational gesture accompanying every clause. This high gestural rate suggests that there might exist more recurrent gesture-speech expressions that are yet to be discovered.

To summarize, given that grammar is rooted in language use, and that gesture is an integral and frequent part of language use (at least in some cases), an adequate constructional approach to grammar must necessarily consider gesture (see Cienki 2015 for the notion of spoken usage event).

2.3 Constructions as building blocks of language in relation to gesture

One basic tenet of construction grammars is that constructions are building blocks of a language: a language is all about constructions and nothing else (Goldberg 1995, 2006; Hilpert 2014). An examination of constructions usually consists of the following aspects: a) how holistic a construction can be; b) how a construction as a whole relates to its parts; c) how the meaning of a construction is situated in human experience (embodied); and furthermore d) how the meaning of a construction is made on the basis of speakers' or hearers' conceptualization of reality. The current section aims to discuss whether, and if so, how, gestural behaviors are compatible with these aspects of constructions.

2.3.1   A holistic view of constructions and gestures

One main attractive point of construction grammars is their holistic view of language. In contrast with generative grammar, this approach holds that syntactic forms and meanings are not in separate, autonomous modules. Instead, grammar consists of symbolic form and meaning pairings. Specifically, the formal side of a construction may involve prosodic information, morphological information, syntactic information, and information about discourse, genre, and/or interaction, etc., while the semantic/functional side of a construction involves semantic, discursive, pragmatic, and/or interactional information. Thus, a comprehensive and holistic description and representation of a construction is provided within this approach, as shown in one example notation diagram in Figure 1:

```
syn    [ external syntactic & categorial properties ]
prag   [ constructional pragmatics & information-structure specifications ]
sem    [ semantics of the construction ]
val    { valence requirements of the construction (set of arguments and adjuncts) }
phon   [ phonological & prosodic properties of the construction ]

role   head                                              role    filler    +
syn    [ head constituent; usually matches external syntax ]    syn ...
prag   [ pragmatics & information-structure of the head ]    prag ...
sem    [ semantic properties of the head ]               sem ...
val    { valence requirements of the head }              etc.
phon   [ phonological & prosodic properties of the head ]
lxm    [ specific lexeme ]
```
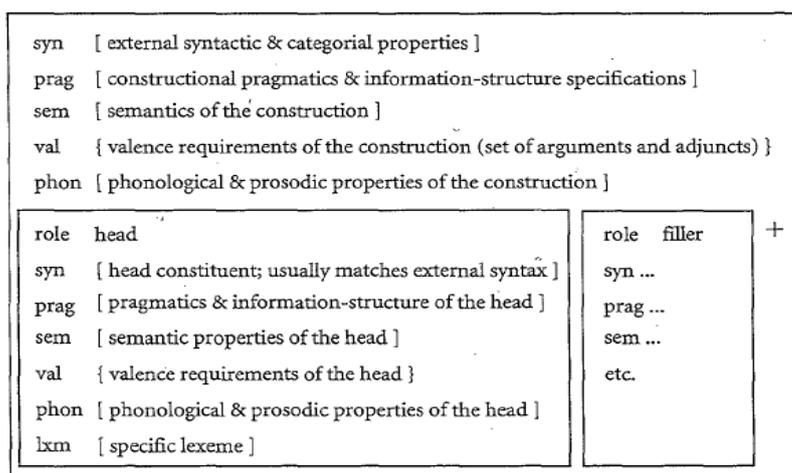
Figure 1 Generic constructional diagram (Fried & Östman 2004: 26)

There are two things to note here. First, a construction is a symbolic unit; that is to say, syntactic forms are inherently meaningful. Second, a grammatical construction, grounded in discourse and context, may involve discourse and contextual information as part of the construction in terms of form and/or function. Properties of gestures seem compatible with these aspects of constructions. More specific discussions are offered below.

a) The symbolic nature of constructions

Construction grammarians hold that forms of constructions are inherently meaningful. Put differently, constructions involve stable form and meaning pairings. This assumption is also present in a variety of gestural studies which concern gestural form and meaning mappings, such as Calbris (1990, 2011) on a list of movement patterns (e.g., curved, looping, circular, etc.), Ladewig (2011, 2014) on the cyclic gesture, Harrison (2008, 2010) on negation gestures, and Kok (2017b) on gestural expressions of uncertainty, prominence, entity or process, et al. (see Bressem & Müller 2014 for a repertoire of German recurrent gestures). As Streeck (2009b: 161) states, Calbris' work is indeed "based on the assumption that gestures – or, more generally, forms (including shapes, Arnheim 1969) – are inherently meaningful". Similarly, Kok and Cienki (2016) also argue in favor of this symbolicity in gesture, proposing that gestural expressions may involve a 'conventional component', which conveys a stable meaning, such as path or motion, and an 'ad-hoc component', which signifies through 'self-symbolization', such as a path derived from a tracing movement. However, as Kok and Cienki have indicated, since a gestural expression (gesticulation) does not "(directly) rely on entrenchment in long-term memory or convention, its symbolic status is of a different nature than that of prototypical

grammatical units" (2016: 81). In other words, it is worth noting that these form-meaning mappings in gesture are still specific to certain verbal contexts (see Kok and Cienki 2016 for more discussion of this question).

Nevertheless, just like constructions in speech, gestural expressions also involve symbolicity, although they have a slightly different symbolic status than prototypical constructions do. In other words, both constructions in speech and gestures involve stativity – stable form-meaning pairings.

b) Constructions functionally situated in discourse and context

Construction grammars hold that grammar is grounded in discourse and context (Du Bois 1987, 2003; Du Bois et al. 2003; Goldberg 2004; Langacker 2001, 2008a). Within this view, syntactic structures of any size, abstracted from usage, may retain any recurrent aspect of the interactive and discourse context as part of their function. Thus, a syntactic pattern of any size may involve the discourse or interactional function, besides the semantic meaning. For example, Du Bois (1987) provides evidence that there is a Preferred Argument Structure in English, involving a preference for expressing lexical and new arguments in subject position in intransitive clauses and pronominal and old arguments in subject position in transitive clauses. This indicates that grammatical and pragmatic dimensions of argument structure can be correlated.

This interactional view of grammar and discourse is also reflected in gestural expressions, in particular in terms of the increasingly recognized multifunctionality of gestural expressions. For example, representational gestures (those depicting referents or events) are also found to relate to a variety of discourse functions, such as marking foregrounded or backgrounded information (Enfield 2004), given or old information (Levy & McNeill 1992; Wilkin & Holler 2011), information as central or peripheral to discourse (Parrill 2010;), verbal disambiguation (Holler & Beattie 2003), or referent tracking (Debreslioska et al. 2013), etc. Furthermore, various dimensions of a single gesture may be linked to different meanings or functions. For instance, hand shapes and dynamic movements of gestures are typically associated with the semantic properties (e.g., entities, process) of events (Kok & Cienki 2016; Wilcox 2004) while gestural space of the same gestures is usually used to signal a discourse function, such as cohesion of discourse (McNeill & Levy 1993; Pereira 2013). These points suggest that representational gestures, which are grounded in discourse and interaction, could retain the recurrent facet of the interactive and/or discourse information as part of their function.

In general, similar to a grammatical construction, gesture could also encode an interaction of grammar and discourse. Specifically, the multifunctionality in gestural

expressions is in line with the situatedness of constructions in speech.

c) Constructions formally situated in discourse and context

A construction may also retain the discourse and interactional context as part of its form. Thus, the form of a single construction may not only involve its internal structure, such as its syntactic properties, but also its external structure, such as discourse context, e.g., constructions of *jestli* (if/whether) in Czech (Fried & Östman 2005), *die Sache ist* (the thing is) (Günthner 2011), and *X-och-X* (roughly means *X and X*) in Swedish (Linell 2009). For instance, *X-och-X* can be used as follows in Swedish,

> (1) (Ongoing talk about a German family that was forced to leave
> Finland after WW II)
> G: sen så beslagtos h<u>u</u>se å dom **flytta** tilb<u>a</u>ka
>   ti ti H<u>a</u>mburg (å)
> M: nå **fl<u>y</u>tta å**[4] **fl<u>y</u>tta** men ja menar va (.) f<u>a</u>n kan du
>   g<u>ö</u>ra
> G: "then the house was confiscated and they **moved** back to to
>   Hamburg (and)"
> M: "well (*nå*), **moved and moved** but I mean what (.) the hell can
>   you do".

As pointed out in Linell (2009), speaker M here seems to accept the sense of 'changing one's living place' and simultaneously to cancel the sense of 'moving voluntarily'. Linell notes that, generally speaking, the 'reactive' construction *X-och-X* is usually used to negotiate the meaning of an expression *X* which occurred in the prior discourse, and to modify the meaning to be accepted locally. Importantly, the construction not only has an internal structure, such as both *X*'s (or at least the second one) being focally stressed, it also has an external syntax, such as this expression usually co-occurring with distancing responsive particles like *nå* ('well') and concessive markers, especially in self-responsive cases, like *i alla fall (*'anyway'), and being followed by a utterance which confirms some aspects and simultaneously cancels some other aspects of X's meaning.

Gestural expressions follow in the same way. Their forms are also sensitive to discourse and context, known as the dynamicity of gesture. First, gesture use is usually sensitive to the speech context, and thus many gestural patterns may retain the speech context as part of their expression, in particular in terms of co-speech gestures (gesticulation). Second, gestural expressions may formally attend to the expressions before or after them, thus resulting in gestural expressions in the form of

---

[4] Note that the *och* 'and' is almost always in the phonological form /o/ in conversation, and thus it is usually rendered as *å* in spelling (Linell 2009).

'catchments' (that is, one or more features occur in at least two gestures in a discourse) (McNeill 2000a). On top of that, gestural forms of one speaker in the interaction may keep the characteristics of another speaker's gestural forms, thus leading to alignment of gestural behavior in different speakers (Holler & Wilkin 2011; Kimbara 2006; Oben & Brône 2016). All these instances indicate that gestural expressions may also retain these interactive and discourse contexts as part of their formal sides.

To sum up, this section has shown that, in line with verbal grammatical constructions, gestural expressions may also to a certain extent be symbolic, as well as situated in discourse or interaction, formally and/or functionally. This holistic view on gesture could thus account for a variety of gestural phenomena, including stativity, multifunctionality, and dynamicity, the latter two of which are usually seen as obstacles for acknowledging gesture's status in grammar.

2.3.2 Compositionality and non-compositionality of constructions and gesture

This section concerns a part-whole relation of a construction, in terms of the (non)compositionality of constructions. It is generally believed that gestural expressions are usually non-compositional (McNeill 1992) while linguistic expressions are usually compositional (in generative grammars). However, constructions can in fact be either compositional or non-compositional in construction grammars, and the same goes for gestural expressions.

a) Non-compositionality

In construction grammars, a construction is viewed as a gestalt, which is not reducible to its components formally and/or functionally. That is, the form and/or meaning of a construction as a whole are/is not compositional from its components. For example, a construction *all of a sudden* cannot be predicted from its components' forms or meanings. Instead, the whole pattern must be learned and stored as a whole. Besides those idiomatic patterns, more productive patterns may also involve non-derivable properties. For instance, a pattern of numeral plus noun, such as *one beer*, is not equal to a sum of its components, since according to a standard syntactic account, an uncountable noun cannot be preceded by numerals.

Gestural expression involves a similar property, known, due to McNeill (1992), as 'global'. The term 'global' means that gesture usually conveys a meaning as a whole and it cannot be reducible to a set of primitives, nor can it be analyzable with compositional rules. For instance, it is almost impossible to derive the meaning of a gesture from its individual form features, such as handedness, hand shape or palm orientation. It is also hard to derive the meaning from the sum of one gestural expression and another in many cases. Instead, the meaning is determined in a

top-down fashion, relying on the context. This is indeed a widely acknowledged characteristic of gesture, and it is also taken as one essential argument against gesture being part of grammar, first claimed by McNeill (1992) and then accepted by many of his followers. However, this feature seems not to constitute an obstacle for recognizing the status of gesture in grammar anymore, if a constructional approach to grammar is adopted; instead, this gestural property can be nicely embraced in construction grammars, as Everett (in preparation) states:

> McNeill's theory, (e.g., 1992, 311ff) takes a perspective similar to Construction Grammar (Goldberg 1995) in claiming that utterances – gesture/speech wholes – are initially 'holophrastic', used as single words or unanalyzable wholes, and later, through reuse and gestural focusing on specific components of the holophrastic construction, analyzed in more detail, leading to grammatical rules in a way reminiscent of the discovery methods of Harris (1947), Longacre (1964), and others, i.e. distributional isolability and recombination.

b) Compositionality

According to the theory of construction grammars, constructions can also be compositional; that is to say, the form and/or meaning of a construction as a whole can sometimes be derived from its components. Thus, if a speaker knows the rules and words, he/she can predict the meaning of the whole expression. For instance, the expression *what time is it* is combined with the syntactic rules of Subject plus Linking verb plus Complement, *wh*-question, inversion, as well as four words – *what*, *time*, *is*, and *it*. In spite of the compositionality of this expression, frequent use of it leads to a conventional form-meaning pairing (Bod 2009).

Indeed, gestural expressions may also involve this compositional property in the sense that (components of) gestures can sometimes be (re)combined. First, in terms of the kinetics in gesture, Fricke (2013) noted that gestural phases can be combined into a larger unit in a recursive way, such as from gestural phases to phrases to larger units. Second, various gestural primitives can be combined to form larger units. Four types of gestural primitives for this are summarized in Kok (2017b: 88-89): a) "lexeme-like gestural patterns such as emblems"; b) "basic phonological parameters", such as hand shape, orientation, location, and movement; c) "morpheme-like patterns such as planes, curves, and lines"; d) "mimetic modes underlying representational mappings". One example of compositionality from Kendon (1996) is as follows. Italian (and also French) speakers extend their thumb and index finger to refer to "telephone", and the speakers move their hands horizontally with palms facing downwards to express cutting off something. When one speaker expresses the

idea of "no one answered the phone", he may move the "telephone" hand, with palm facing down rapidly to the right. Thus, a gesture referring to "telephone" is combined with one expressing "cut off", forming a larger unit.

To summarize, construction grammars, which can account for both compositional and non-compositional patterns in language, are well equipped to cover the complexities of gestures – either as global or as compositional.

2.3.3   Meaning as embodied, inside and outside

An embodied view of the meaning of a construction is adopted in construction grammars and also in cognitive linguistics in general (Barsalou 1999; Bergen & Chang 2005; Johnson 1987; Lakoff & Johnson 1980, 1999). It is assumed that linguistic meaning is grounded in the human body in that when speakers/hearers comprehend or produce utterances, sensory-motor affordances of these utterances are evoked in speakers'/hearers' minds. This assumption has been supported by converging evidence from various approaches, such as behavioral, neural, and psychological studies (Glenberg & Kaschak 2002; Stanfield & Zwaan 2001; Pulvermüller 1999; Zwaan et al. 2002). For instance, Stanfield and Zwaan demonstrated that people create a mental simulation of the orientation of an object implied by a sentence. Participants were faster to respond to pictures with the orientation of an object implied by the sentence they read than pictures which did not match (2001).

Mental simulation and gesture have been claimed to be tightly coupled, exerting mutual influences on each other (Marghetis & Bergen 2014). First, speakers' mental simulations seem to shape their own gestures. For instance, gesture is claimed to arise from simulated action and perception, with the likelihood of gesture production rising as the degree of mentally simulated action increases (Hostetter & Alibali 2008; Hostetter & Alibali 2010). Second, speakers' gesturing could shape their own mental simulation. For instance, representational gestures have been claimed to help speakers to maintain mental imagery (de Ruiter 2000), or to help the conceptual planning of motor-spatial information for speech production (Kita 2000), or to evoke motor-spatial information (Beilock & Goldin-Meadow 2010). Next, speakers' gestures could shape listeners' simulation. For example, viewing gestures with details of action can lead listeners to include those details in their subsequent re-production of the action (Cook & Tanenhaus 2009); Wu and Coulson (2007) found that, after viewing clips involving utterances accompanied by gestures (e.g., *it's actually a double door* + a gesture for Dutch-style double doors, one part above the other), participants found it easier to respond to pictures in which speech and gestures match (that is, pictures showing Dutch-style double doors) than to those in which they do not match (e.g., pictures showing French-style double doors, the two parts

being side by side). This indicates that viewing gestures may lead listeners to create detailed visual information.

Drawing upon the above close relations between mental simulation and gesture, Marghetis and Bergen (2014) propose that mental simulation and gesture, as forms of internal and external embodiment, are tightly coupled. That is, meaning-making is tied to both mental simulation and gesture:

> Meaning is embodied internally when we create embodied simulations, co-opting brain areas specialized for perception or action to create dynamic mental representations, rich in sensorimotor detail… At the same time, meaning is embodied externally in representational gestures, actions of the hands and body that represent objects, actions, and ideas. (Marghetis & Bergen 2014: 2000)

Given that the meaning of constructions is tied to both mental simulation and gesture, it stands to reason that a multimodal approach to constructions is necessary.

2.3.4   Meaning as conceptualization, visible in speech and gesture

Many construction grammarians and cognitive linguists in general hold that a linguistic structure reflects a speaker's construal of an event, rather than simply the objective reality (Croft 2012; Goldberg 1995, 2006; Langacker 1987, 2008a). In other words, speakers' choices of grammatical structures in speech (such as transitive vs. intransitive, perfective vs. imperfective aspect, and nouns vs. verbs, etc.) reflect their different conceptualization of reality. For instance, to express an event whereby someone exerted some force on a vase and the vase broke, a speaker may say *he/she broke the vase* or *the vase broke* or other expressions, depending on what she/he wants to profile at that moment. By using the former clause, the speaker profiles both causation and the change of state it brings about; by using the latter, he/she only profiles the change of state (Langacker 1990, 2008a). In general, the following major dimensions of conceptualization associated with constructions are distinguished by Langacker (2008a): specificity (level of precision), focusing (arrangement of conceptual content in terms of foreground and background) and prominence (focusing of attention, related to focusing), and perspective (viewing arrangement).

Initial studies show that these dimensions of conceptualization can also be reflected in gestural expressions (see Cienki 2015; Kok & Cienki 2016; Wilcox 2004). First, the gestures speakers employ can display the specificity of speakers' conceptualizing of a referent, such as in terms of a more schematic image schema (like PATH) (Cienki 2005) or a more specific 'mimetic schema' (like RUN) (Zlatev 2005).

Second, speakers could also use gesture to foreground and background information (prominence and focus), such as in terms of the asymmetry of hands (Enfield 2004) or in different types of pointing gesture – either B-pointing (usually "big" in form, e.g., full arm and eye gaze often aligned) or S-pointing (usually "small" in form, e.g., hand only and "casual articulation") (Enfield et al. 2007), etc. Furthermore, a speaker's perspective on an event (e.g., an internal or external perspective) can also be displayed in gesture, such as in terms of Character Viewpoint gesture or Observer Viewpoint gesture (McNeill 1992), or in terms of more or less complex gestural forms (Duncan 2002), etc.

The above discussion suggests that how speakers construe an event is not only reflected in their choice of linguistic structures but also in their choice of gestural representation, such as gestural forms or depicting methods (refer to Cienki 2015, Kok & Cienki 2016, and Wilcox 2004 for a more detailed discussion on this question).

All in all, this section has demonstrated that various gestural phenomena are in line with various aspects of constructions, including symbolic status, interaction of grammar and discourse, (non)compositionality, the embodied nature of meaning, as well as meaning being equal to conceptualization.

2.4 The lexicon-syntax continuum and gestures

Another basic tenet of construction grammars is that there is no clear-cut boundary between lexicon and syntax. Instead, all levels of grammar can be represented as constructions, from morphemes, words, phrases to sentences. What differs among constructions is only their size/complexity and schematicity, as shown in Table 2.1.

Table 2.1 The syntax-lexicon continuum (Croft 2001: 17)

Construction type	Traditional name	Examples
Complex and (mostly) schematic	Syntax	[SBJ be-TNS VERB-en by OBL]
Complex and (mostly) specific	Idiom	[pull-TNS NP-'s leg]
Complex but bound	Morphology	[NOUN-s], [VER-TNS]
Atomic and schematic	Syntactic category	[DEM], [ADJ]
Atomic and specific	Word/lexicon	[this], [green]

Gestural expressions might also be distinguished along these two dimensions of size/complexity and schematicity.

First, gestural expressions differ in the degree of schematicity, as discussed in Cienki (2015) and Kok & Cienki (2016). Kok and Cienki suggest that various types of

gestures distinguished by McNeill (1992, 2005) differ in their degree of schematicity, increasing from emblem (e.g., the OK sign), pointing, recurrent gestures (e.g., circular movement of hands or fingers referring to continuity of an activity) to gesticulation (that is, spontaneous co-speech gestures). If pantomime is included, and if gesticulation referring to 'mimetic schemas' (e.g., putting something in a container, or running) is distinguished from gesticulation referring to 'image schemas' (e.g., tracing movement of one forefinger referring to the path or motion of an entity), then the previous continuum needs to be revised. The gestural categories of emblem, pantomimes, and gesticulation referring to 'mimetic schemas' seem to be more specific and more identifiable formally and/or functionally, whereas the gestural categories of pointing, recurrent gestures, and gesticulation referring to 'image schemas' appear to be more schematic, more context-dependent, and less identifiable.

Second, these gestures of various categories seem to differ in their size as well, increasing from emblems and pointing gestures to the other categories above. The former two types of gesture seem to be more word-like, since they usually function as particular words such as *OK, this, that, here*. In contrast, the other categories, including recurrent gestures, pantomime of events, and gesticulation, seem to be more complex and more phrase-or-sentence-like, since they seem to function as a group of words rather than single words. Considered together in terms of these two dimensions, Table 2.2 proposes a continuum for gestures of the various categories mentioned above.

Table 2.2 A continuum of various types of gesture

Gesture type	Traditional name	Examples
Complex and (more) schematic	Recurrent gesture; Gesticulation referring to 'image schemas'	Cyclic gesture (circular movement of hands or fingers referring to continuity of an activity); Tracing movement of one forefinger referring to the path or motion of an entity
Complex and (more) specific	Pantomime of events; gesticulation referring to 'mimetic schemas'	Manual movement in which a speaker holds a cup hand shape and moves the hand towards a point at the speaker's front, as if he/she put a cup-shaped object at the front; Body movement in which a speaker moves as if to run.
Atomic and more schematic	Pointing	Pointing at a point in a space with a forefinger, thumb or flat hand
Atomic and more specific	Emblem	The "Okay" emblem gesture

2.5 The taxonomic organization of constructions and gestures

Another basic tenet in construction grammars is that constructions are organized in a network according to various inheritance relations. Goldberg (1995) distinguishes four major types of inheritance links, including the polysemy link, the subpart link, the instance link, and the metaphorical link. First, a polysemy link refers to a link between the central sense and the extensions of a construction. Goldberg (1995) holds that each of the extensions can be seen as a minimally different construction, motivated by the central sense. One example for this would be various senses of the ditransitive construction: "X CAUSES Y to RECEIVE Z" (central sense), e.g., *Joe gave Sally the ball*; conditions of satisfaction imply "X CAUSES Y to RECEIVE Z", e.g., *Joe promised Bob a car*; "X ENABLES Y to RECEIVE Z", e.g., *Joe permitted Chris an apple*. Second, a subpart link involves one construction being part of another, such as the intransitive motion construction, which can be seen as part of the caused motion construction, e.g., *a ball rolled out of the stadium* and *she rolled the ball out of the stadium*. Next, an instance link exists when one construction is a specific case of another. For instance, the lexical item *drive* can instantiate the resultative construction, such as in *Chris drove Pat crazy*. Last, a metaphorical link is posited when one construction relates to another one via a metaphorical mapping, such as the resultative construction (e.g., *Pat hammered the metal flat*), which seems to be a metaphorical extension of the caused motion construction (e.g., *Pat threw the piece of metal off the table*). The mapping between the two is that the resultative construction seems to encode a metaphorical change of location while the caused motion construction encodes a physical change of location.

The above types of links among constructions seem to exist among gestural expressions as well. First, a polysemy link above seems to exist among gestural expressions which share similar morphological forms but involve various meanings, such as in the recurrent gesture families. For instance, the cyclic gesture formally involves "continuous circular outward movement"; semantically it has a core of "cyclic continuity" and various meanings, such as ongoing events or ongoing mental activities (Ladewig 2011, 2014). In other words, these meanings are interrelated with each other, all of them being motivated by the semantic core.[5]

Second, a subpart link could be proposed for the phenomenon whereby one gestural pattern can be seen as part of another one. Various gestural modes of representation distinguished by Müller seem to be interrelated with each other in this way, in particular in terms of the combined modes. Specifically, a) Embodying

---

[5] However, to my knowledge, there has been little discussion on which sense is the central one and which are the extensions, in the case of this recurrent gestural family.

gestures[6] can be understood as a subpart of Tracing-with-Embodying gestures. For instance, speakers would use one stretched forefinger to embody a long and thin object (e.g., a microphone) while they would use one stretched forefinger with a change of location to represent the movement of this object (e.g., movement of a microphone). It can be seen that the former gesture is part of the latter one formally and semantically; b) Tracing gestures (those tracing trajectory movements of entities, rather than tracing outlines of entities, e.g., a gesture depicting the movement of a door) can be seen as part of Acting ones (e.g., a gesture depicting the activity of someone opening the door, involving the movement of the door), since the latter gestures usually embed an internal movement, which Tracing gestures represent (images of such Tracing and Acting gestures to depict the door-opening activities can be found in Chapter 6).

Furthermore, an instance link could connect specific gestures with more schematic ones. To begin with, representational gestures of the same depicting mode may differ in their specificity. For example, Acting-with-specified-object gestures (e.g., representing the activity of putting a cup on the desk) appear to be less schematic instances of Acting-with-unspecified-object ones (e.g., representing the activity of putting something on the desk); similarly, Tracing-with-specified-object gestures can be understood as instances of those with an unspecified object. On top of these, representational gestures may also be linked to the related schematized pragmatic gestures in this way. For instance, in the case of a representational presenting gesture (e.g., presenting a medal to a winner) and a pragmatic presentational gesture, known as the Palm-Up-Open-Hand gesture (PUOH) by Kendon (2004) and Müller (2004), the two can be similar in terms of their movement pattern, but the former is usually more specific than the latter, both formally (in terms of the hand shapes) and semantically (in terms of what is represented in relation to speech with the hands and movements) (see also Chapter 8 for examples of representational presenting gestures). The former can thus be understood as an instance of the latter.

Metaphorical links among various gestural patterns, however, are a rather complicated issue, since all gestures are by nature literal movements. Consequently, this will not be discussed here.

The discussions in this section indicate that inheritance links among constructions are helpful in providing a way to consider the relations among various gestural patterns. In this way, an organized network of various gestural expressions

---

[6] Embodying gestures refer to gestures belonging to the Embodying mode. The same goes for Acting gestures (i.e. gestures belonging to the Acting mode), Tracing gestures (i.e. gestures belonging to the Tracing mode), and Molding gestures (i.e. gestures belonging to the Molding mode), et al.

could possibly be established in the future.

2.6 Summary and possible ways to research multimodal constructions

This chapter has demonstrated that a usage-based approach to constructions in construction grammars requires an examination of all aspects in language use, and an adequate construction grammar definitely needs to include gesture as part of its overall model. In addition, the main features of constructions are compatible with the properties of gestures in terms of the basic tenets in construction grammars, including gestural properties which have been taken as crucial arguments against admitting their status in grammar, such as dynamicity, multifunctionality, and non-compositionality/unpredictability. This consistency between gesture and grammatical constructions makes it possible to incorporate gestures as part of grammar. All these considerations indicate that it is not only necessary, but also possible, to consider the multimodality of constructions.

Following from the aforementioned basic tenets of construction grammars, there could be the following possible ways to investigate the multimodality of constructions.[7]

1) Conventional, multimodal form-meaning pairings

Construction grammars maintain that conventional form-meaning pairings – constructions –are building blocks of the language system. It is thus a key and unavoidable issue to identify stable multimodal form-meaning pairings, i.e. conventional form-meaning gestural pairings accompanying certain verbal constructions.

2) Dimensions of construal in constructions and gestures

Another important tenet in construction grammars is that linguistic structures reflect how we conceptualize events in the world. A question that follows is whether, and/or to what extent, various dimensions of event construal, afforded by various constructions, could be expressed in gestural representations. An affirmative answer to this question can be seen as providing support for the "growth point" hypothesis (McNeill & Duncan 2000) as well as the multimodal nature of grammar (see also Cienki 2016; Kok & Cienki 2016; Lapaire 2011).

3) (Non)Compositionality of multimodal constructions

A further important property of grammar is its ability to combine semantic and/or formal primitives. Thus, construction grammarians have been endeavoring to discover how individual components combine into a whole construction. Similarly, this can be another area for multimodal studies in construction grammars, which

---

[7] It is also possible to investigate monomodal gestural constructions following from these tenets.

could focus on how gestures and verbal constructions combine into multimodal constructions, or what the semantic contributions of gestures and verbal constructions are, respectively, in multimodal constructions.

4) Intersection of grammar and discourse in multimodal constructions

As discussed above, construction grammars admit an intersection between semantics and pragmatics, by holding that grammatical constructions may involve both semantic and pragmatic information. It is thus worth examining how multimodal constructions are situated in both grammar and discourse. One research route in this regard could be about the multifunctionality of multimodal constructions; that is, to discover the semantic and pragmatic aspects of a gestural pattern accompanying a certain verbal construction. Another route might be to determine how an intersection point of grammar and discourse found in verbal constructions in previous work, such as the preferred argument structure found in conversational discourse (Du Bois 2003), is co-expressed in the accompanying gestures.

5) Schematicity of multimodal constructions

The schematicity of multimodal constructions could provide a fruitful avenue for future research. Since constructions are essentially grounded in the body and selective in their encodings, it is necessary to identify the mechanism for this selection process. This issue has indeed constituted a separate, individual area of research in cognitive linguistics, in particular in work on image schemas. (For discussions on the multimodality of image schemas, see Cienki 2005, 2013c.)

6) Inheritance links among multimodal constructions

As mentioned above, the language system is an organized network, which connects grammatical constructions via various inheritance links. It is worthwhile to examine how multimodal constructions link with each other, for example, how a multimodal caused-motion construction links with a multimodal intransitive motion construction, how a multimodal caused motion construction relates to a multimodal resultative construction, or how a central multimodal ditransitive construction (e.g., *Joe gave Mary a watch*) connects extended multimodal ditransitive constructions (e.g., *Joe promised Mary a watch*). Researching this question could help us to obtain a better understanding and characterization of the organization network of multimodal constructions.

7) Multimodal constructions in other fields

A further direction would be an application of multimodal constructions in related fields, such as in terms of how children acquire multimodal constructions, how language learners learn multimodal constructions, how teachers teach multimodal constructions, how multimodal constructions are represented in the human brain,

and many others. Note that these issues do not directly concern the present thesis, although some of them have already been investigated (see for example Andrén 2010, 2014 on the acquisition of multimodal constructions).

It is clear that there are a number of possible directions that multimodal research in construction grammars could take. To date, studies in this field have mainly focused on the first three issues, including multimodal expressions as conventional pairings, conceptualization of events encoded in constructions and gestures, and the compositionality of gestures and constructions in multimodal constructions, rather than the issues concerning the interaction between grammar and discourse,[8] or the inheritance relations among multimodal constructions. This is understandable, given that the field of multimodal construction studies is still young and has been focusing on the basic, internal structures of multimodal constructions, rather than the similarities/differences between or links among multimodal constructions (that is, various constructions on a lexicon-syntax continuum or the inheritance links among constructions). The next chapter will provide a review of what has been done in terms of these topics. For discussions on the schematicity of multimodal constructions and the application of multimodal constructions in related fields, readers are referred to others areas in cognitive linguistics (such as works on image schemas in Cienki 2005, 2013c, and the field of applied cognitive linguistics in Goldin-Meadow et al. 2001, Goldin-Meadow & Singer 2003; Gullberg 2008, 2013; Stam 2008, 2010, 2015), and thus they will not be reviewed in this thesis.

---

[8] Although the multi-functionality of gestures has received some initial attention (e.g., Kok et al. 2016), an intersection of semantics and discourse functions of specific multimodal constructions has largely been unexplored.