Chapter 3. Previous studies on three prerequisites of multimodal constructions

Although there has been an increasing interest in the multimodality of constructions, existing studies have particularly focused on a number of particular perspectives, as mentioned at the end of the previous chapter. More specifically, they include a) to what extent co-expressions of gestures and verbal constructions constitute conventional multimodal constructions; b) to what extent the means of conceptualization of events, which are visible in the encoding of verbal constructions, can be reflected in the accompanying gestures; and c) how gestures and verbal constructions combine into multimodal constructions. All of these perspectives can be seen as prerequisites for a multimodal construction (see also Zima & Bergs 2017; Hoffman 2017).

In this chapter, I first give an introduction to previous studies on the multimodality of constructions from these perspectives. Then I consider some issues that are not addressed in these studies, which give rise to the research questions handled in the present thesis.

3.1 Conventional, multimodal constructions

A major interest in the multimodality of constructions in construction grammars has been trying to identify conventional multimodal form-meaning pairings (that is, gestural patterns + verbal patterns). A common means to this end has been to examine whether a verbal construction is frequently accompanied by a certain gestural pattern, either adopting a gesture-first approach (that is, gesture as the starting point of analysis) (Bressem & Müller 2017) or a verbal-pattern-first approach (that is, verbal patterns as the starting point of analysis) (Lanwer 2017; Mittelberg 2017; Cánovas & Valenzuela 2017; Schoonjans 2014; Zima 2014, 2107). Usage-based construction grammars hold that if a certain pattern is frequently used, it should be considered as a construction (Bybee 2006; Bybee & Hopper 2001; Goldberg 2003, 2006; Langacker 2001, 2008a). Similarly, if a verbal construction frequently co-occurs with a particular gesture, the recurrent multimodal pattern should be considered as a multimodal construction. Apparently, in these studies, the "frequency" of gestures accompanying verbal constructions is a key criterion to define multimodal constructions. To date, this perspective has attracted the most empirical studies and brought many fruitful results.

To begin with, some studies have discovered a number of conventional multimodal patterns on a phrasal or clausal level, such as Zima (2014, 2017), and

Cánovas and Valenzuela (2017). As one of the few initial empirical studies on particular multimodal constructions, Zima (2014), employing a large multimodal database – the UCLA Library Broadcast NewsScape (the database which was also used in this dissertation, as will be introduced later) –, found that the following three constructions were frequently accompanied by particular gestures in American English: [V(motion) in circles], [N (3rd pers sg) spins around], and [all the way from X Prep Y]. For instance, when speakers used the pattern [all the way from X Prep Y] in speech, 76% of the time they made gestures to depict a path or a distance. Cánovas and Valenzuela (2017) found that a phrasal construction – the [from X to Y] construction (e.g., *from beginning to end*), which is associated with the timeline of an event – was often accompanied by a certain spatial gestural pattern (8 out of 10 instances). Given the frequent co-occurrence of verbal constructions and the accompanying gestures which can depict similar meanings, these constructions are claimed to be multimodal constructions.

Discourse-level constructions are another area of interest for studies on the multimodality of constructions, such as Bressem and Müller (2017) on constructions expressing negative assessment in German, Schoonjans (2014) on discourse particles in German, Lanwer (2017) on patterns of narrow and loose appositions in German, and Jehoul et al. (2017) on expressions showing obviousness in Dutch. For instance, Schoonjans (2014) found that gestures and some discourse markers in German frequently co-express certain discourse functions. Specifically, when German speakers used the discourse marker *einfach* (“only”, “simply”), they not only performed a Palm-Up-Open-Hand gesture (PUOH) but also shaked their heads 69.72% of the time as they carried out these gestures. According to Schoonjans, these frequently used multimodal expressions should be seen as multimodal constructions. Importantly, Schoonjans pointed out that these co-expressions can best be considered as single multimodal constructions, rather than two individual monomodal constructions – gestural constructions and verbal constructions seperately – due to the fact that verbal and gestural elements are optional to each other’s formal and functional characterization. This is an initial study which made an explict argument in favor of multimodal constructions and against monomodal ones.

It is more remarkable that abstract grammatical constructions are also found to be frequently accompanied by relevant gestures. Hinnell (2014) found that all of the five aspect-marking patterns she examined – [start/stop/quit/keep/continue + V-ing/To V] – were frequently accompanied by relevant gestures, which reflect the meanings of these aspects. For instance, gestures accompanying the patterns [start/stop/quit + Ving/To V] were found to involve fewer gestural phases than those accompanying the patterns [keep/continue + Ving/To V]. Moreover, gestures were

found to be frequent in terms of all syntactic patterns, although the pattern [continue + To V] was accompanied by most gestures – 75% – whereas the pattern [start + Ving/To V] was accmpanied by fewest gestures – 50%.

While the above studies mostly focused on specific constructions in relation to gestures and used absolute frequencies of these gestures, Kok (2017a) took a top-down approach to co-expressions of gesture and speech and also considered the relative frequencies of gestures accompanying a certain construction in a corpus. Specifically, on the one hand, he investigated all gestures accompanying any words and parts of speech in a multimodal German corpus of route descriptions; on the other hand, he applied collostructional analysis (see Stefanowitsch & Gries 2003 for this approach) to co-expressions of gesture and speech, by considering frequencies of gestures accompanying certain words or parts of speech as well as frequencies of gestures accompanying all other words or parts of speech in the whole corpus (this is called the relative frequency of gestures accompanying a certain construction). Kok found that certain words tend to be positively related with gestures, which he called "gesture-attracted words", while some other words tend to be negatively related with gestures, which he labelled "gesture-repellent words". Prototypical examples of "gesture-attracted words" include deictic expressions such as *hier* ("here"), discourse particles such as *quasi* (so to speak), and spatial words such as *rund* ("around"), whereas prototypical examples of "gesture-repellent words" include non-spatial words such as *glauben* ("to believe") and pronouns. As for parts of speech in relation to the accompanying gestures, he found that pronominal adverbs, nouns, determiners, prepositions, and adverbs are gesture-attracted whereas injections, pronouns, filled pauses, participles, and other word classes are gesture-repellent. That is to say, certain words and parts of speech have a greater potential to be part of multimodal constructions than others do. Kok's study provides a sense of words and parts of speech which have greatest potential to be part of multimodal constructions. In addition, as for the methodology, this approach is relatively more objective than the others mentioned above, since a consideration of the relative frequency of gesture precludes the possibility that the likelihood of gestures accompanying a certain verbal construction is simply due to chance.

The empirical studies above indicate that gestures could frequently occur with verbal constructions on various levels of the language system, ranging from words and phrases to discourse and abstract lexical or syntactic categories. All these provide a critical challenge for usage-based approaches to constructions, which aim to pursue all recurrent aspects in language use in theory but simply focus on the texts or speech in language use in practice. Instead, these frequent gestural and verbal co-expressions of various levels offer important empirical insights into the

potential to acknowledge the existence of conventional multimodal constructions to some extent.

It is worth pointing out a theoretical question which arises in these studies, that is, the issue of frequency, as noted in Cienki (2017b), Hoffman (2017), Schoonjans (2014, 2017), and Zima and Bergs (2017). As shown in the above empirical studies, the rate of gestures accompanying a certain construction can be higher (e.g. approximately 70%) or lower (e.g. around 50%). Construction grammars hold that if the use of a pattern is frequent enough, it should be considered as a construction. A question follows as to how frequent is 'frequent enough' to be a construction; to put it differently, what frequency constitutes sufficient entrenchment of a multimodal pattern. This question, Schoonjans (2017) points out that is not unique to multimodal studies, as it also arises in "traditional" constructional studies. This issue thus does not constitute a substantial obstacle to multimodal studies. Although Schoonjans points out that the frequency issue should not deter us from exploring the multimodality of constructions, he does not directly offer a substantial answer to the question of how to deal with multimodal patterns with different gestural frequencies. Motivated by the fact that gesture is often an optional component, and that gesture can be more or less frequent with a dynamic scope of relevant behaviors (e.g., hand, head, or shoulder movement) for many verbal expressions, depending on contextual factors, Cienki (2015, 2017b) pushes the discussion a step further, by proposing the concept of "variable multimodality". He suggests that the multimodality of a construction should be viewed in terms of more or less multimodal, rather than of "multimodal" or "not multimodal". Following this idea, the right question to be asked and researched should be to what extent co-expressions of gestures and verbal constructions are entrenched multimodal constructions (more or less), rather than what the cut-off point is for the gestural rate in defining an entrenched multimodal construction. This idea constitutes one of the bases of this thesis.

3.2 Means of event construal in constructions and gestures

Another way to approach the multimodality of grammatical constructions is to investigate whether various dimensions of construal, reflected in the choice of constructions (or syntactic encodings), can be manifested in the use of the accompanying gestures. If the answer is found to be positive, it suggests that gestures and verbal constructions might be controlled by the same conceptual structure (similar to the idea of "growth point" in McNeill & Duncan 2000); that is to say, there might exist a multimodal construction from which both gestures and verbal constructions derive. The major difference between this approach and the

above one in Section 3.1 is that the present one mostly deals with two or more related constructions with different means of event construal, rather than individual constructions. Studies from this perspective will now be discussed in more detail.

Initial and widespread gestural studies from this perspective include those investigating gestures in relation to motion constructions (Chui 2012; Duncan 2001; Furman 2012; Kita & Özyürek 2003; Kita et al. 2007; Özyürek & Kita 1999; Özyürek et al. 2005).[9] The general aim of these studies is to ascertain whether different linguistic encodings of the motion event in different languages, which are associated with different ways in which speakers conceptualize the events according to the "thinking-for-speaking" hypothesis (Slobin 1987, 1996), are accompanied by different gestures. Gestures are found to differ in ways in which events are packaged and conceptualized in respective languages, which is seen as lending support to the proposal that gesture interplays with grammatical constructions, as predicted by the "thinking-for-speaking and gesturing" hypothesis (McNeil & Duncan 2000) and the Interface hypothesis (Kita & Özyürek 2003). For instance, through examining how American English, Turkish, and Japanese speakers gesture with respect to motion events, Kita and Özyürek (2003) found that the Path in motion events was less likely to be gestured by Japanese and Turkish speakers than by English speakers, which is associated with the hypothesis that this information is more difficult to verbalize in Japanese and Turkish than in English. The same study also found that Japanese and Turkish speakers were more likely to make two different gestures to refer to the Manner[10] and the Path respectively, whereas English speakers were more likely to make one gesture to refer to the two features of the motion event. This is claimed by the authors to be associated with the syntactic property that English packages these two features more tightly, in terms of morphological combination, than Turkish and Japanese do. In general, these studies have provided preliminary evidence for a relation between gestures and syntactic encoding, although most of them do not explicitly employ the framework of construction grammars, except for Furman (2012). However, each syntactic encoding of the motion event in different languages may encode a habitual conceptualization rather than a dynamic online conceptualization of the event during speaking, as pointed out in Kita et al. (2007). It follows that it remains unclear whether gestures are influenced by the habitual conceptualization of events or their online conceptualization.

In order to overcome this problem, Kita et al. (2007) furthermore tested whether different syntactic encodings of Manner or Path in motion events within the

[9] There are only a few studies on placement events (Gullberg 2011), location events (Tutton 2011), and others.

[10] "Manner" refers to the specific way that a figure moves (e.g., walking).

same language (that is, in English in that study) co-occurred with different gestural representations. In Kita et al's study, participants were asked to watch clips of the same motion event and then to retell the event with either two clauses or one clause. Analysis of gestures accompanying these clauses in speech showed that different syntactic encodings were often accompanied by different gestures. More specifically, when speakers used one clause (e.g. *he rolled down the hill*) in speech, they often performed one single gesture which integrated both Manner and Path components of this motion event; when speakers used two clauses (e.g. *he went down as he spun*) in speech, they tended to perform two gestures referring to Manner and Path components separately. This study indicates that gestures might interact with online syntactic encodings of events rather than the habitual conceptualization of events formed within a certain language, by considering constructions within the same language. As a first study to investigate the relation between gestures and syntactic encodings in a language, Kita et al.'s study provides further support for a relation between gesture and grammar.

It is worth noting that both types of studies above, within the English language or across languages, addressed the relation between gestures and the means of construal afforded by motion constructions, which controlled the type of motion events involved. By doing so, they preclude the possibility that the relations between gestures and the means of construal are simply reducible to the relations between gestures and the type of events.

A different line of research has focused on abstract grammatical categories – aspectual categories – in relation to the accompanying gestures, either within a certain language or across languages, such as in English, Chinese, French, German, and Russian (Becker et al. 2011; Boutet et al. 2016; Cienki & Iriskhanova in press; Duncan 2002; McNeil 2003; Parrill et al. 2013; Wang 2017). These studies are grounded in a hypothesis similar to that in the above studies, namely that different grammatical categories are associated with different means by which speakers conceptualize events and thus are expected to be expressed in different forms in gestures. The earliest study on this topic is Duncan (2002), which examined gestures in relation to progressive and non-progressive aspects in Chinese and English. It was found that gestures accompanying progressive aspectual constructions tend to last longer in their strokes and involve more complex forms than those accompanying non-progressive aspectual constructions. Later, Becker et al. (2011) examined a different set of aspectual categories in English – Aktionsarten categories (states, activities, achievements, and accomplishments) – in relation to gesture. It was found that gestures with achievement verbs tended to have a punctual ending whereas those with activity verbs tended to have an extended or repeated movement.

Building on these studies, a project was carried out (Cienki & Iriskhanova, in press) to examine this topic on the basis of French, German, and Russian, since these languages may involve interestingly different forms of thinking-for-speaking according to the thinking-for-speaking hypothesis (Slobin 1987, 1996). In the study reported in Cienki and Iriskhanova (in press), the three languages displayed different patterns of interrelation between the (im)perfectivity of verbs and the (un)boundedness (with or without ballistic pulse of effort) of gestures. Specifically, a significant correlation was found between the use of the French perfect tense form (the *passé composé*) and bounded gestures, and between the imperfect forms (the *imparfait*) and unbounded gestures, a difference which was not found with the perfect(ive) versus imperfect(ive) tense and aspect forms in the German and Russian data. All these studies indicate that the means of construal afforded by grammatical constructions seem to be reflected in different ways in the accompanying gestures across languages. This furthermore suggests that gestures are in various ways part of the "linguistic-conceptual representation" of abstract aspectual constructions, rather than simply the pre-linguistic imagistic properties of events.

3.3 Compositional multimodal constructions

Just like "frequency" and "event construal" discussed above, compositionality, also referred to as "code integration" (Fricke 2013), has been taken as another essential criterion to define a multimodal construction (see also Zima & Bergs 2017). A compositional multimodal construction basically refers to a multimodal pattern in which both a gestural pattern and a verbal pattern semantically and/or structurally contribute to a multimodal expression. Ziem (2017) proposes using the following deletion test to assess the compositionality: If the gesture is deleted, and the meaning of the construction becomes uninterpretable, then this multimodal expression is called a multimodal construction. Note that this phenomenon usually occurs with deictic expressions, which only account for a limited proportion of the elements of the language system. It follows that only limited empirical studies have followed from this tradition.

Fricke (2013) is an early study on defining multimodal constructions in this regard, [11] drawing upon the multimodal contribution to German noun phrases. In general, Fricke observed that when German speakers used *son* ("such a"), they usually made a gesture to complement the relevant information. More specifically, if a deictic gesture (e.g., a speaker pointed at a real table in front of her/him) accompanied the word *so* ("such a"), as in the utterance *Ich will sonen Tisch kaufen*

[11] Although Fricke (2013) was based on the Eisenberg's grammar (1999), the discussion seems also to be applicable to construction grammars.

("I want to buy such a table"), the co-expression of gesture and verb referred to the type of "table" which was similar to the one she/he pointed at, but the properties of the table were not expressed; that is to say, this multimodal expression refers to an entity with a definite type but an indefinite token. If a representational gesture (e.g., a speaker made a gesture to refer to the shape and size of the "table") accompanied the word *so,* as in terms of the same utterance *Ich will sonen Tisch kaufen* ("I want to buy such a table"), the co-expression of gesture and speech depicted the specific properties of the entity, but the type was not mentioned; in other words, this multimodal expression has a definite token but an indefinite type. In both cases, the expression of noun phrases consists of both gesture and speech, whereby the noun in speech functions as the head of the noun phrase and the gesture fills in as a modifier. Accordingly, this is considered as a multimodal construction, in which neither gesture nor speech can be absent. This challenges the prevalent view of verbal constructions as primary in linguistics proper.

Gesture can not only complement the information provided by noun phrases, but can also substitute for syntactic constituents of clauses, as shown in Ladewig (2012). Specifically, Ladewig found that gestures preferably take over the functions of noun phrases or verb phrases in final clausal positions. For instance, while a speaker said *Ich wollte dieses* ("I wanted this"), the speaker made a gesture, which started from the word "this" and referred to the "wall" of a dam. This can be seen as supplementing the objects in the utterance. In this case, gestures and verbal clausal constructions are combined into a multimodal construction, in which both modalities seem to be equally important.

Despite the limited multimodal compositional constructions found, the above studies indicate that it is worth including the gestural modality as part of linguistics proper, given the obligatoriness of both verbal and gestural expressions. As an important criterion for defining a multimodal construction, the use of gesture deserves further empirical study to reveal more about this topic.

3.4 Shortcomings of existing research and motivations for the present thesis

We have seen that a number of studies have explored the issue of multimodality of constructions and yielded a number of results in the past decade. Existing studies, however, are still not sufficient in a number of respects, some of which I will now deal with.

First, as for the conventional form-meaning pairings, although the multimodality of constructions on various linguistic levels has been touched upon, the multimodality of clausal-level constructions has largely been left unexplored. Given

that clausal constructions are basic units that speakers use to build a discourse (Langacker 2008a: 354), without studies on these constructions, a discussion on the multimodality of constructions would be far from complete. In addition, a few studies which have tried to do so have paid attention to constructions expressing dynamic motion events, especially with typological studies, as discussed above. As is indicated in Thompson and Hopper (2001), these constructions only account for a small proportion in conversations. Thus, the multimodality of many other constructions, which might be used more frequently than motion constructions are, remains unknown to us. Against this background, this thesis aims to examine the multimodality of clausal constructions by taking the most basic and frequently used clausal constructions in spoken language (in this case, English) as the starting point.

Second, empirical studies on the dimensions of construal in relation to gesture within one language (as noted above) are still not sufficient. Most attention from this perspective has been given to aspectual categories, with the exception of Kita et al. (2007). Given that this grammatical category is only a small part of a language system, it remains unclear as to whether a relation between grammar (dimensions of construal in this aspect) and gesture exists in the other grammatical constructions, such as the above-mentioned basic clausal constructions. Therefore, this thesis will investigate to what extent the kinds of event construal associated with these basic, frequently used constructions can be reflected in the accompanying gestures.

In addition, the semantics of constructions resides in the interaction between the construal and the conceptual content in a frame, the two aspects of which cannot be isolated from each other. It is thus necessary to ascertain whether gesture correlates with the dimensions of event construal or simply the types of events/frames. However, few gestural studies have tried to distinguish between the two aspects except Kita et al. (2007), which addressed the relation between gesture and constructions depicting motion events, as discussed above. According to construction grammars, these two constructions in Kita and colleagues' work encode two distinct forms – one clause and two clauses – and accordingly, encode obviously different online conceptualizations. It is, thus, debatable as to whether gestures interact with constructions which involve relatively more similar syntactic forms and more closely related conceptualizations than the previous syntactic encodings do, such as the dative alternation (e.g., *she gave me a book* vs. *she gave the book to me*), the locative alternation (e.g., *Jack sprayed paint on the wall* vs. *Jack sprayed the wall with paint*), and others. The thesis will address this question by considering constructions with the same frames and closely related forms (that is, within the single clauses or with the same type of transitivity), which I will refer to as constructional alternations. This is expected to provide further insight into a relation

between gesture use and the means of event construal associated with constructions.

It is worth pointing out that previous research on gesture and grammatical constructions has predominantly concentrated on representational gestures and has largely ignored discourse-related gestures. For instance, most gesture models have focused on the origin of representational gestures, such as the "Interface Hypothesis" of Kita and Özyürek (2003), and Hostetter and Alibali's (2008) framework of Gestures as Simulated Action. It thus remains a mystery as to whether representational gestures and discourse-related gestures derive from the same cognitive origin or not, or whether discourse-related gestures correlate with grammar in speech or not, although there is a so-called 'general activation' conjecture, which predicts that events or referents with a strong action component not only activate representational gestures but also prime gesturing in general (Masson-Carro et al. 2016). Therefore, the present study aims to provide some insights into this issue by including both representational gestures and discourse-related gestures in the research. Next, as for (non)compositionality, although unpredictability — another dimension of the compositionality of constructions — is an important critierion for defining a construction in traditional construction grammars, to my knowledge, no studies have tried to define a multimodal construction in this way. That is, no studies have tried to investigate to what extent the form or meaning/function of a multimodal construction cannot be predicted from its components, which in theory could involve the following aspects: 1) a multimodal construction as a whole conveys some meaning, which is not equal to the sum of verbal meaning and gestural meanings in this construction, or which is not predictable from the meanings of other recognized multimodal constructions; and/or 2) the form of a multimodal construction as a whole cannot be predicted from the syntax of speech and/or the rules of gesture,[12] or from the forms of other recognized multimodal constructions. Given that gestalt understanding is an important cognitive ability, pursuing this way of understanding a multimodal construction would provide more insight into the nature of multimodal constructions. However, this thesis will not go into this question for reasons of time and space and, more importantly, due to the fact that the previous two questions concern more basic and urgent issues in pursuing the multimodality of constructions.

[12] By "rules" of gestures, I mean the principles that govern behavior of a certain gesture or a sequence of gestures (such as the recursion rule of gesture found in Fricke 2013 or any other rules to be found in the future studies) or the common patterns in which a certain gesture or a sequence of gestures are produced; but the issue of rules of gesture is still quite challenging.

Given the above motivations, this thesis focuses on the issues related to the first two prerequisites of multimodal constructions, including the issues of stable form-meaning pairings and dimensions of construal in constructions in relation to gesture (including representational, deictic, and discourse-related gestures), rather than the (non)compositionality of multimodal construction candidates. Specifically, it aims to investigate the following research questions:

a) Is there a stable gestural pattern with respect to each of the basic and frequent clausal constructions?

b) To what extent does this gestural pattern reflect the means of construal afforded by these constructions?

c) How often does each gestural pattern occur with each of these grammatical constructions, or to what degree are the co-expressions of gesture and verbal constructions conventionalized multimodal constructions.

To approach these questions, the following constructions are considered: a) a group of basic and frequently used constructions in spoken language, and b) constructional alternations, as mentioned above. In addition, different groups of constructional alternations are further included, driven by the following rationale in general. As will be introduced in the coming chapter, constructions in various constructional alternations may involve the same or different scopes of predication (that is, the "onstage region"), although they all could evoke the same semantic frame. For instance, constructions in the causative alternation, such as *she opened the door* vs. *the door opened*, involve different scopes of predication; constructions in the dative alternation, such as s*he gave the book to me* & s*he gave me a book,* involve the same scope of predication. Distinguishing between the above two types of constructional alternations could help to discover to what extent gestural use is sensitive to the means of event construal afforded by various constructions in various contexts rather than to properties of the frames activated. In doing so, a close relation between gesture and the means of conceptualization, if there is one, might be revealed.

In the coming chapter, I will give an introduction to the basic clausal constructions and constructional alternations to be investigated in this thesis.