Alan Cienki*

Utterance Construction Grammar (UCxG) and the variable multimodality of constructions

Abstract: Some proponents of the theory of Construction Grammar have been investigating how it might address the nature of spoken language usage as multimodal. Problems confronted in this endeavour include the variability with which gesture is used with speech in terms of its (in)frequency and its (non)obligatoriness: for some expressions a certain kind of gesture is basically obligatory, but for most others it is a variably optional component depending on contextual factors. This article proposes "utterance" as a level of description above that of speech and gesture for characterizing audio-visual communicative constructions. It picks up on earlier proposals to consider constructions as prototype categories with more central and more peripheral features. The language community's knowledge of a given utterance construction and that of any language user are discussed as "deep structures" (in a non-Chomskian sense) that provide a set of options (some more central and others more peripheral) for expression, whereby any "surface structure" is a metonymic precipitation in context of the construction's features. An important attentional mechanism proposed that guides production and comprehension ("uptake") of utterance constructions is the dynamic scope of relevant behaviors. Taken together, this approach may help bring Construction Grammar closer to being a truly usage-based theory.

Keywords: Construction Grammar, gesture, utterance, prototype category

1 Multimodal Construction Grammar?

In some ways, recasting the theory of Construction Grammar (CxG) (Fillmore et al. 1988; Goldberg 1995; and others) to account for talk as multimodal is an appealing prospect. The openness of the approach toward exploring patterned connections between form and meaning/function lends itself to including gestural expressions to the degree that they, too, are part of form-meaning/function pairings that are sufficiently entrenched and conventionalized. The potential for a multimodal version of construction grammar has already been proposed by a number of scholars (e. g., Andrén 2010; Schoonjans 2014; Steen and Turner 2013; Zima 2014).[1] However, to make the maximal claim that the grammar of any given language is 100 % multimodal (in the sense of audio-visual) can be seen as problematic. For example, speakers continue to use audio-only radios and telephones as a means of long-distance communication, and even when in each other's presence, listeners do not always have speakers in view. Hearers understand speakers' talk broadcast on radio or by telephone, and of course the assumption is that a person speaking or writing is not using a completely different form of the grammar of a given language, but that at most, one is a variant of the other. (Although see more on the telephone issue in Section 5 below.)

A number of theoretical questions arise when we envisage CxG taking the multimodal turn. Some of them are issues that are already being debated by monomodally-focussed construction grammarians, but which come to the fore with greater saliency when considering the possible role of other modalities or modes in the theoretical framework. For example, the issue of what constitutes sufficient entrenchment of a

---

1 It is worth clarifying here that *multimodal* will be used in this article as a cover term for the use of multiple semiotic modes or codes as well as for the combination of audible and visible communicative actions, which technically would better be called bimodal communication as involving two sensory modalities, but for which the customary term is multimodal communication.

---

*Corresponding author: Alan Cienki, Language, Literature and Communication, Vrije Universiteit Amsterdam, Amsterdam, Netherlands; Moscow State Linguistic University, Moscow, Russian Federation, E-mail: a.cienki@vu.nl

lexico-grammatical form-meaning pairing for it to be considered a construction gains an added level of complexity when one adds the question: What constitutes sufficient entrenchment of gestures with a verbal construction for them to be counted as part of the construction? Furthermore, can we speak of gestural constructions that may be monomodal, not occurring with verbalizations? The theoretical questions are complicated by the fact that gesture form-function pairings vary in the degree to which they are codified and have normative standards of production in a given culture. For example, for *emblems* (Efron 1941/1972), the form-meaning/function pairing is more fixed; these are gestures that function as conventional signs within a given culture, such as the French gesture of screwing the tip of one's forefinger into one's temple to indicate that some is crazy (Calbris 1990: 4). However, for so-called *recurrent gestures* (Bressem and Müller 2014b; Ladewig 2014; Müller 2004), the pairing is between a family of forms and a family of functions (thus Kendon 2004 uses the term *gesture families*). An example would be the family of *away gestures* (Bressem and Müller 2014a) used in German and many other cultures, whereby the hand moves in the horizontal plane away from the body, indicating some kind of rejection or negative assessment, with subcategories including holding away, brushing away, throwing away, and sweeping away. At the end of this gesture continuum are idiosyncratically produced gestures for which the relation between form and meaning is much more context-dependent and may involve one-time production of gestures, especially for depicting specific referents (Müller et al. 2013).

2 Utterance Construction Grammar (UCxG)

Is there a way to sensibly treat the array of complexity constituting verbo-gestural expression within the framework of CxG? This article presents one proposal, building on some existing approaches in construction grammar and gesture studies. The proposal starts from the idea of considering the *utterance* as the entry point for characterizing constructions, rather than assuming that gestures should be plugged in to existing verbally-based frameworks, or conversely, assuming that gestures form the higher level of constructional structure, and words fit into them. The latter has affinities with the claim of Armstrong et al. (1995) that speech is (oral) gesturing that produces sound, and thus spoken language falls under the higher-level category of gesture. However, given the normatively conventionalized system of verbal language, and the fact that in most face-to-face interactions (between hearing-seeing adults in contemporary industrialized societies) gesture is on the whole more dependent on speech for its communicative function than speech is on gesture (vid. Feyereisen et al. 1988; Kibrik and Molchanova 2013), using gesture as the starting point for a CxG seems counter-intuitive for a construction grammar model of utterances in their usual manifestations.

There have been other approaches to characterize an *utterance grammar*. Weigand (2010) is one example. A point in common between that proposal and the one below is the "aim to explain how speakers succeed in establishing coherence among the different communicative means offered to them in the action game" (p. 88), with the "action game" being Weigand's development of Wittgenstein's (1953) idea of *language game*. However, the approach proposed here does not adhere to Weigand's theoretical reliance on speech act theory nor to the assumptions that go with it about the nature of linguistic categories. Indeed, Weigand is opposed to cognitive and functional theories of grammar whose starting point is characterized as "the expression side" of language, and so she pushes them aside as "pragmatic grammars" (pp. 90–91). The approach espoused in the present article is also unrelated to the use of the term *utterance grammar* in computational linguistics, where it can refer to attempts to formalize dialogic talk (e. g., Core and Schubert 1996; Dowding et al. 1993).

What I am proposing here with Utterance Construction Grammar is inspired by the goal of thinking about constructions in multimodal terms (or rather, as potentially involving multiple modalities), rather than in terms of purely verbally expressed lexico-grammatical entities. Following Kendon (2004: 7), "we shall use the term 'utterance' to refer to any ensemble of action that counts for others as an attempt by the

actor to 'give' information of some sort". Kendon notes that he is drawing upon Goffman's (1963: 13–14) distinction between action via which people may "give off" information about themselves to others who are co-present (e. g., about their character, mood, or social status) versus action "that is regarded as explicitly designed for the provision of information and for which they [the producers] are normally held responsible:" (Kendon 2004: 7). He sums it up that *utterance* can be considered "the ensemble of actions, whether composed of speech alone, of visible action alone, or of a combination of the two, that counts for participants as a 'turn' or 'contribution' or 'move' within the occasion of interaction in which they are engaged" (Kendon 2004: 110). The term *gesture* will thus usually be used in this article to cover what Kendon (2015: 44) characterizes as "utterance dedicated visible bodily action" and *speech* or *spoken language* will be used to refer to "utterance dedicated audible bodily action". Taking *utterance*, in all its complexity, as the starting point for theorizing about constructions is in line with Langacker's broad approach to considering what forms of expression may become conventionalized as part of a linguistic sign in Cognitive Grammar. According to Langacker (2008: 457), the phonological pole of a usage event "includes the full phonetic detail of an utterance, as well as any other kinds of signals, such as gestures and body language".

3 Deep and surface structures, UCxG style

The acronym UCxG is a purposeful play on the UG of Universal Grammar from the tradition of generative linguistics. Though UCxG is being used here without the nativist assumptions that underlie generative UG theory, ironically, UCxG might be characterized in terms of a deep structure and a surface structure, albeit it with a different theoretical view on what deep and surface pertain to than in traditional UG theory. In our approach, the deep structure of a construction can be seen as a set of tools that can be drawn upon to express the construction (surface structure). The surface structure is thus a metonymic representation of some (if not all) elements of the construction. We will return to this issue in Section 5 below.

First, however, we should distinguish two ways of thinking about the deep structure and the surface structure in UCxG. One is in terms of what a linguist/grammarian might describe as an Utterance Construction (UCx) in all its potential detail. This would be like the level of analysis currently used in CxG, but would also incorporate any potential gestural (or other) elements that may be used in some instantiations of the construction. Let us call this theoretical deep structure of an UCx its *DeepS-potential*. One distinction here is that DeepS-potential does not have to do with generic forms of UCxs or with more schematic versions of them; rather, the DeepS-potential of an UCx is a theoretical construct that includes any and all elements that may occur in any behavioral realization of the construction. These are the elements that could be stored cognitively as playing a role in a given UCx. It is important to note here that UCxs are understood here as having a prototype structure; therefore, the question is not whether an element is or is not part of an UCx, but rather how central or peripheral it is to that UCx, or more specifically, to the DeepS-potential of that UCx. The prototype-nature of UCxs is discussed further in Section 4 below. In addition, the range of language users to whom the DeepS-potential of an UCx is claimed to pertain is dependent upon the scope of the language community about which the linguist is making claims (a particular subset of the population of a particular city? monolingual speakers of a given language in a particular site? etc.).

The surface structure potential or *SurfS-potential* will be any potential instantiation of an utterance construction, whether attested or not. This category includes instantiations posited by linguists – otherwise known as invented examples. If DeepS-potential is the -emic level of an UCx (analogous to phoneme or morpheme), then SurfS-potential is the level of all the possible allomorphs of the UCx.

A second way of thinking about a DeepS in UCxG is in terms of a model of the knowledge of an UCx by a given language user at a given moment in time – the knowledge of the UCx that s/he has "stored" (at his/

her cognitive disposal) at a given time. Let us call this deep structure of an UCx the *DeepS-user*. In turn, *SurfS-used* will refer to any actual documented (e. g., video-recorded) instance of usage of a given utterance construction. From another perspective, SurfSs-used can be seen as *constructs*, in the sense of Fried (2015: 980) – concrete utterance tokens that have actually been produced – that, through processes of entrenchment (Langacker 1987: 59; Schmid 2007), can come to be categorized into DeepSs-user by those acquiring the language.

The simple term *DeepS* will be used to refer to deep structure in terms relevant to both of the above characterizations of it. Similarly, *SurfS* will be a cover term for the two ways of thinking about surface structure noted above.

4 Utterance constructions as prototype categories

Following on proposals by Gries (2003), Imo (2007), and Lakoff (1987: Case Study 3) and others for verbally-based models of CxG, any UCx should be seen as having a prototype structure, which can be thought of as a center-periphery model in terms of the status of the elements or features that comprise it. The term *features* will be used here sometimes to refer to the component elements of a construction; however, the features here are not to be associated with notions of "markedness theory", the "semantic features" of formal semantics, or the like. The proposal is that features comprising a DeepS often do not have equal status, but rather some features are more central for the construction than others (more prototypical, in the sense of corresponding to the central tendency of the category as a whole [Barsalou 1987: 104]). Similarly, some features expressed in the SurfS-potential may be judged by members of a linguistic community to be more prototypical of a given construction, and some features in a given SurfS-used in a given context may be judged by listeners/viewers as having been more key for recognizing the instantiation of the UCx (having higher cue validity, in the sense of Rosch and Mervis 1975). (At this stage, these are proposals that await empirical testing.) Prototypicality here should therefore be seen as a factor of salience that makes a given UCx easily recognizable to language users – either on average across a community or in a given instance of use (a *usage event*, to use the term of Langacker 1988 and elsewhere) – rather than necessarily being a feature that occurs more frequently with the UCx. Now let us get a sense of what might constitute the various features in light of multimodal communication (or more precisely: variably multimodal communication). We will do so by simply considering UCxs on the meta-level for the moment (without distinguishing deep versus surface levels or status as *potential* or as *used*). The examples below are all drawn from North American English, unless noted otherwise.

As previous research noted above has suggested, various kinds of expressive forms (verbal, gestural, intonational, etc.) may play a role in constructions. For some UCxs, the lexico-grammatical elements may be the more prototypical ones and constitute the driving force behind the constructions. For these, any potential gestural features would be less prototypical and less predictable in character. For example, in a current study (Wu, In Preparation) using the UCLA Library Broadcast NewsScape, accessed via the Distributed Little Red Hen Lab (redhenlab.org), copular clauses with evaluation (e. g., "That's nice"; "That's good"; "I'm sure") have been found to be infrequently accompanied by gestures, as compared to clauses of other types (such as transitive clauses). Note the related results of Kok (2016), who found verbs of cognition in his German data (*wissen* 'to know' and *glauben* 'to believe') to be particularly "gesture repellent". This can be contrasted with the situation with deictic constructions, where gesture typically plays a more central role: the use of gestures with, or even in place of, deictic referential constructions is well known (going back at least to the Wittgenstein 1953 discussion of ostensive definition). For example, flat-hand pointing gestures, with the extended fingers together pointing laterally, for indicating a route direction that someone should follow (Kendon 2004: ch. 11) are a good example of when gesture can play the leading role (e. g., in answer to a question about a local destination that begins with, "How do I get to ... ?"). In many cases, the gesture alone can suffice as one's utterance.

The shrug and "I don't know"/"I dunno" to express not knowing something or one's distanced stance is a notorious example of a behavioral compound (Streeck 2009: 189 ff.). Both words and gestures share central status in the UCx, but the internal structure of the construction's SurfS-potential is flexible: its various elements can serve as central cues of this UCx, either individually or in combination, momentarily leaving other cues with peripheral status. In terms of gestures, features used may include shoulder lifts (one or both shoulders), palm-up open-hand(s), a puckered mouth, possibly a head tilt to the side. The verbal elements of the utterance can range from a full phrase "I don't know" to a reduced "I dunno", and even a low-high-mid intonation contour can suffice for the vocalized element. The various individual features or combinations of them can serve as cues of this UCx with different semantic/pragmatic nuances per combination, for example: with shoulder lifts in particular being able to independently carry the specific function of acknowledging a stance differential (with the interlocutor or with a previously stated position), but with a full-fledged verbo-gestural performance of "I don't know", with a complete upper-body shrug, more emphatically indicating one's epistemic state of lacking knowledge and affective state of disaffiliating oneself from the situation (Debras and Cienki 2012). The multiple possible realizations of the DeepS as different, but interrelated, SurfSs make this UCx a good example of what can be called an UCx family (analogous to the gesture families, mentioned above) (see Croft 2007 on the organization of constructions in networks).

For other UCxs, the various features share central status individual component features but are less likely to be pushed to the periphery in SurfS realizations. Consider the presentation of oneself, or of one's accomplishment, with a theatrical "ta daa!", exclaimed in a high, steady pitch with the second syllable extended and with both one's arms spread out wide, hands open. For such an UCx, the compound of elements work together, often as a whole-body performance, to make up a recognizable instantiation of the UCx. A potential SurfS (i. e., a given SurfS-potential) could metonymically only express one feature of the complete DeepS-potential (e. g., for this UCx, using just the words spoken in a low volume, or just one arm spread outward). However, for such UCxs in which features appear to share central status as a gestalt ensemble, I would predict that it would be harder for listening viewers to pick up on the use of single-features as instantiations of the UCx than it would be for them to interpret single elements of the shrug/"I don't know" behavioral compound as SurfS instantiations of that UCx.

The sample types of UCxs noted above suggest that the model of UCxG can provide a solution to the conundrum of how to handle multimodality in construction grammar in at least three ways. (1) It considers constructions as complexes whose structure is rooted at the level of the utterance, rather than at a modality-specific level of speech or gesture. This avoids the necessity of (artificially) having to decide if a given construction is a verbal one that has a gestural component or whether it is a gestural construction with a verbal component (consider the "I don't know" UCx, described above, where the decision could go either way). (2) It views the Cx as having a center-periphery structure in which the potential component elements have differing statuses as more or less prototypical of the construction. This is a more flexible alternative than positing that the model has the binary choice between required and optional elements. (3) The prototype model handles the conundrum of *sufficiency* of entrenchment of a form-meaning/function pairing needed in order to "qualify" as a construction: the present approach views symbolic pairings as more or less entrenched, thus making the status of a given "construction" a matter of degree; some pairings constitute more stable, recognizable constructions than others.

The model assumes that UCx are dynamic, on various, overlapping time scales. A given language user's knowledge of an UCx will vary according to his/her experience in the language community. A peripheral element can be used saliently by one person in a given context and therefore be picked up initially by another person as a central feature of that UCx. For example, in developmental terms, Crystal (1979) discusses the initial use of appropriate intonational tunes by children 14–16 months old in interactional sequences before they are able to utter the phrases that customarily go with them. He cites (p. 39) the example of a child using extended high and mid tones that the parent used when saying "all gone" at the end of a meal and only being able to reliably produce the segmental phonemes for the utterance about a month later. Dore (1975) argues that prosodic features of children's holophrases can be seen as the building

blocks of their acquisition of speech acts as wholes, which later become grammaticalized syntactic structures.

Gesture can also play a more salient role initially in constructional use within first language acquisition. Acredolo and Goodwyn (1985) discuss how children between 12.5 and 17.5 months can produce symbolic gestures before being able to utter coordinate expressions verbally. This applies not only to object labels (e. g., two fingers opening and closing for 'scissors') but also to expressing not knowing something (shoulder shrug with hands palm up). Namy et al. (2004) discuss the U-shaped curve in 18-month old children's use of gestures for naming, followed by a decrease in gesture use (age 26 months) as they acquire words as the primary means of symbolic reference, followed by a later (4 years old) acceptance of gestural symbols alongside their use of speech. These developmental changes in the use of different constructional features can be accommodated by a prototype model that allows for dynamic movement within the constructional category (DeepS-user) from periphery to center and/or vice versa.

An important factor here is what behavioral cues are being deemed communicatively relevant for the expression of the meaning/function of a given utterance construction at a given moment in time – and over time, across contexts, with the stabilization of a construction's symbolic structure. This is the issue taken up in the next section.

5 Dynamic scope of relevant behaviors

Let us return to the view of the SurfS as being a metonymic expression of some features of the DeepS. Two different producers of communication from the same language community may share essentially the same knowledge of an UCx (their DeepSs-user of the UCx may be very similar), but they may, in a given usage event or even customarily, use different SurfSs of that UCx. Each individual may only draw upon some elements of the DeepS in their SurfS expression to invoke the UCx. This is the sense in which SurfS expression can function metonymically (or to be more precise: synecdochically, with a part standing for the whole). In fact, such *pars pro toto* functioning of constructional expression may usually be the norm.

Note, however, that ellipsis of features in a SurfS-used does not necessarily mean that they are missing in the DeepS-user. Indeed, a culturally normalized form of ellipsis via metonymic expression of part(s) of the construction (cf. Barcelona 2005) could mean that some features of a construction's DeepS are rarely produced in SurfSs-used,[2] perhaps only in contexts of elaborated repairs of misunderstood utterances. While prototypical features of an UCx may be more prototypically selected for expression in SurfSs, in theory any selection of components of the DeepS can appear in the SurfS. As Mark Turner (personal communication, 11 March 2016) phrased it, "the product is a given precipitation of a process." Whereas more prototypical features may be more likely to be expressed, and picked up on, by listeners/viewers as cueing a given UCx, the final result in a given usage event depends on many contextual factors, and the model predicts a range of degrees of salience/obviousness (from greater to lesser) of instantiation of a given UCx, depending on the cues used. Consequently, there will be a range of degrees of intelligibility by listeners/viewers, depending on the selection and combination of cues used and the context of the usage event. How this plays out with specific multimodal UCxs is a field of study still in its early stages, something to which we hope this special issue of this journal will contribute.

---

2 Thanks to Bertie Kaal for this inspiration. Note, also, that the issue here is akin to that of "null instantiation" of elements of a construction (e. g., Fried and Östman 2004: 66–69), but also different from it; rather than purely being about the filling in or not of syntactic elements, here we are dealing with the optionality of both sequential and simultaneous forms of expression in potentially multimodal UCxs. Issues of not only complementarity but also redundancy of expression as natural qualities of variably multimodal DeepSs require further explication in this model. See Cienki (In press: Lecture 2) for initial discussion of these issues.

A key point for this proposed model of UCxG is that in usage events of communication, interlocutors employ and take into consideration a dynamic scope of relevant behaviors (see Cienki 2012, 2015a, 2015b, 2015c for details). Rather than referring to the tradional terms *speaker* and *addressee* (or listener/viewer, as above), I will henceforth use the broader categories of *producers* (to theoretically encompass speakers, people gesturing when not speaking, and people using a sign language) and *attenders* (those paying attention to someone else's communication). Building on work from Relevance Theory (Sperber and Wilson 1986), the selective activation of meaning (Müller 2008; Müller and Tag 2010), and attentional analysis of meaning (Oakley 2009), the claim (Cienki 2012: 155) is that a given producer's focus on what behaviors to deploy as relevant in a given context, and an attender's focus on which behaviors of the producer are relevant in a given context, are both variable; this means that the producer's and attender's scope can be broader or narrower, involving more or fewer behaviors (e. g., speech+gesture+facial expressions in an emotional context versus just the words spoken in a fact-oriented recitation of items). Presumably when the size of both the producer's and attender's scopes are aligned with each other, their communication should be more in sync. But the scope of relevant behaviors is dynamic not only in terms of how it can zoom in or out, but also in that its focus can shift. Whereas for speakers, spoken language in face-to-face interaction arguably constitutes the default focus, the focus can move temporarily from this gravitational center to other behaviors, such as gesture (such as when one is trying to communicate through a soundproof glass wall).

Consider the example of a telephone call in which each participant may be doing some gesturing while speaking, but each cannot see the other's gestures. Though it may seem to be an example of monomodal, purely audio communication, that is not necessarily the case for the speaker at any given moment, who may be seeing (or at least feeling, via proprioception) the gestures s/he is producing. The scope of relevant behaviors can thus differ (and probably usually does differ) for producer and attender at any given moment. Furthermore, the production of gestures while speaking can have effects on the prosody of co-produced speech (Krahmer and Swerts 2007), resulting in multimodal effects for the hearer as well if we consider *modal* here as referring to modes or semiotic codes of communication (in this case: lexico-grammatical items and prosodic patterns).

In terms of the model of UCxG introduced above, greater incorporation of different elements from different modalities in a given SurfS-used of an UCx can be described in terms of a producer's use of a broader scope of relevant behaviors of the features that constitute his/her DeepS-user of the UCx. Different UCxs may have developed from scopes of relevant behaviors of different breadths; some constructional DeepSs strongly bias production of a SurfS with a broader scope of relevant behaviors (like with the presentational "ta daa!" example cited above). Others DeepSs offer a wide variety of options for expression (like that of the "I dunno-shrug" UCx), while others entail a more minimal range of options for expression (the name of a rare chemical formula can be written as a compound with the abbreviations of its component elements, or named in words in spoken language, but it probably does not have a conventional gesture for its expression among most language users). In turn, a given attender (call her *Anne*) may "have" a DeepS-user of an UCx that incorporates a wide scope of behaviors as relevant to it (especially if Anne is considered a masterful speaker/performer), but her interlocutor (call him *Bob*) may produce only a minimal realization of that UCx based on the meager DeepS-user of that UCx that he is in command of (perhaps Bob is a more word-oriented person who picks up less on others' co-verbal behaviors). The resultant SurfS-used of Bob as producer (e. g., a monotone "I agree" with minimal body movement) may be more difficult for Anne as attender to comprehend if she is expecting a more elaborate performance of the UCx, entailing as relevant a broader scope of behaviors (e. g., an "I agree" with enthusiastic intonation, raised eyebrows, and a head nod). In this sense, attenders like Anne may need to rely on more contextualization cues (Gumperz 1982) as a way to make sense out of "what's going on here?" with a given utterance in context if the scope of relevant behaviors used by people like Bob is much less then she might expect, based upon her DeepS-user (and the DeepS-potential of it that other speakers share with her) which offers a broad scope of options for expression.

6 In closing

Utterances, in all their complexity, are proposed here as the starting point for analyzing constructions. They constitute a point of access to spoken, gestured, or signed data that is neutral with regard to the verbal/gestural distinction. Indeed, such an approach is consistent with Kendon's category of utterance-dedicated perceptible bodily action as one encompassing gesture, speech, and sign.

Deep and surface structures have been proposed in a new guise as a way to tease apart the potential form features of a construction (either theoretically, as DeepS-potential, or as hypothesized knowledge of a given language user in DeepS-user) from how the construction is realized (either in its theoretical possibilities, in SurfS-potential, or in attested usage events, as SurfS-used). SurfSs are metonymic realizations of DeepSs, with full realization of all features of a DeepS being the limiting case. The distinction of (1) -potential versus (2) -used/-user is a way to characterize (1) the level of the system and possible realizations of that system from that of (2) individuals' hypothesized knowledge of the system and their actual implementation of it in usage events of communication.

Prototype models of categories provide a way to handle the differential internal structure of constructions and different degrees of entrenchment of features. This also affords explanations about the variability in the degree to which features of constructions may be subject to being lost or replaced; more central members would be more resilient to being lost than more peripheral ones. But the prototype models should be seen as dynamic to accommodate change, such as movement of peripheral features to more central positions in the UCx through entrenchment or normative pressure in the linguistic community.

Finally, the dynamic scope of relevant behaviors was invoked as an attention mechanism that is presumed to be engaged in any communicative usage event, both in the process of the producer's mobilizing selected communicative behaviors in the moment, and in the process of the attender's dynamically setting the range of his/her "communicative radar" to pick up on (or not) features of UCxs employed by producer. The range of each of the producer's and attender's scopes varies dynamically and the focus of each also shifts dynamically in response to cognitive load and attentional factors as well as contextual factors. Consequently, the features of a DeepS-user that are realized in a SurfS-used may differ per producer from one usage event to another, which accounts for variability in production of a given UCx. In addition, the features of the UCxs picked up by an attender may be a subset of those produced by someone else, but presumably communication is more effective when producers' and attenders' DeepSs are more similar to each other's, and when the scope of features selected as relevant in context for expression of a UCx by a producer is close to the scope of behaviors that the attender is focussing on as relevant.

It is hoped that these proposals can help circumvent some of the problems confronted when trying to imagine a multimodal form of construction grammar. Admittedly, they only offer a schematic introduction to the relevant ideas, and in that sense they are themselves a mere gesture to the reading audience for consideration.

References

Acredolo, Linda P. & Susan W. Goodwyn. 1985. Symbolic gesturing in language development. *Human Development* 28. 40–49.
Andrén, Mats. 2010. Children's gestures from 18 to 30 months. Sweden: Lunds Universitet Unpublished PhD dissertation.
Armstrong, David F, William C. Stokoe & Sherman E. Wilcox. 1995. *Gesture and the nature of language*. Cambridge: Cambridge University Press.

Barcelona, Antonio. 2005. The multilevel operation of metonymy in grammar and discourse, with particular attention to metonymic chains. In Francisco J. Ruiz De Mendoza Ibáñez & M. Sandra Peña Cerval (eds.), *Cognitive linguistics: Internal dynamics and interdisciplinary interaction*, 313–352. Berlin: Mouton de Gruyter.

Barsalou, Lawrence D. 1987. The instability of graded structure: Implications for the nature of concepts. In Ulrich Neisser (ed.), *Concepts and conceptual development*, 101–140. Cambridge: Cambridge University Press.

Bressem, Jana & Cornelia Müller. 2014a. The family of away gestures: Negation, refusal, and negative assessment. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Jana Bressem (eds.), *Body – language – communication. An international handbook on multimodality in human interaction*, 1592–1605. Berlin: De Gruyter Mouton.

Bressem, Jana & Cornelia Müller. 2014b. A repertoire of German recurrent gestures with pragmatic functions. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Jana Bressem (eds.), *Body – language – communication. An international handbook on multimodality in human interaction*, 1575–1592. Berlin: De Gruyter Mouton.

Calbris, Geneviève. 1990. *The semiotics of French gesture*. Bloomington, IN: Indiana University Press.

Cienki, Alan. 2012. Usage events of spoken language and the symbolic units we (may) abstract from them. In Janusz Badio & Krzysztof Kosecki (eds.), *Cognitive processes in language*, 149–158. Bern: Peter Lang.

Cienki, Alan. 2015a. The dynamic scope of relevant behaviors in talk: A perspective from cognitive linguistics. In *Proceedings of the 2nd European and the 5th Nordic Symposium on Multimodal Communication*, Tartu, Estonia, August 6–8, 2014. http://www.ep.liu.se/ecp_article/index.en.aspx?issue=110;article=002.

Cienki, Alan. 2015b. The notion of the dynamic scope of relevant behaviors in cognitive linguistic theory. In Andej A. Kibrik & Aleksej D. Koshelev (eds.), *Language and thought: Contemporary cognitive linguistics*, 560–573. Moscow: Languages of Slavic Culture. [Ченки, Алан. 2015b. Понятие динамического диапазона коммуникативных действий в теории когнитивной лингвистики. Ред. Андрей А. Кибрик и Алексей Д. Кошелев. *Язык и мысль: Современная когнитивная лингвистика*, 560–573. Москва: Языки славянской культуры].

Cienki, Alan. 2015c. Repetitions in view of talk as variably multimodal. *Vestnik of Moscow State Linguistic University* 6(717). 625–634.

Cienki, Alan. In Press. *Ten lectures on spoken language and gesture from the perspective of cognitive linguistics: Issues of dynamicity and multimodality*. Leiden: Brill.

Core, Mark G. & Lenhart K. Schubert. 1996. Dialog parsing in the TRAIN system. *Technical Report 612*. March 1996. Rochester, NY: The University of Rochester.

Croft, William. 2007. Construction grammar. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 463–508. Oxford: Oxford University Press.

Crystal, David. 1979. Prosodic development. In Paul Fletcher & Michael Garman (eds.), *Language acquisition*, 33–48. Cambridge: Cambridge University Press.

Debras, Camille & Alan Cienki. 2012. Some uses of head tilts and shoulder shrugs during human interaction, and their relation to stancetaking. In *Proceedings of the ASE/IEEE International Conference on Social Computing*, 932–937.

Dore, John. 1975. Holophrases, speech acts and language universals. *Journal of Child Language* 2(1). 21–40.

Dowding, John, Jean Mark Gawron, Doug Appelt, John Bear, Lynn Cherny, Robert Moore & Douglas Moran. 1993. Gemini: A natural language system for spoken-language understanding. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, 54–61.

Efron, David. 1941. *Gesture and environment [= 1972. Gesture, race and culture]*. New York: King's Crown Press.

Feyereisen, Pierre, Michèle Van De Wiele & Fabienne Dubois. 1988. The meaning of gestures: What can be understood without speech? *Cahiers de Psychologie Cognitive/European Bulletin of Cognitive Psychology* 8(1). 3–25.

Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticiy in grammatical constructions: The case of *let alone*. *Language* 64(3). 501–538.

Fried, Mirjam. 2015. Construction grammar. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – theory and analysis: An international handbook*, vol. 2, 974–1003. Berlin: De Gruyter Mouton.

Fried, Mirjam & Jan-Ola Östman. 2004. Construction grammar: A thumbnail sketch. In Mirjam Fried & Jan-Ola Östman (eds.), *Construction grammar in a cross-language perspective*, 11–86. Amsterdam: John Benjamins.

Goffman, Erving. 1963. *Behavior in public places*. New York: Free Press.

Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Gries, Stefan T. 2003. Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1. 1–27.

Gumperz, John J. 1982. *Discourse strategies*. Cambridge: Cambridge University Press.

Imo, Wolfgang. 2007. Der Zwang zur Kategorienbildung: Probleme der Anwendung der *Construction Grammar* bei der Analyse gesprochener Sprache. Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion 8. 22–45. www.gespraechsforschung-ozs.de.

Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kendon, Adam. 2015. Gesture and sign: Utterance uses of visible bodily action. In Keith Allen (ed.), *The Routledge handbook of linguistics*, 33–46. London: Routledge.

Kibrik, Andrej A. & Natalia B. Molchanova. 2013. Channels of multimodal communication: Relative contributions to discourse understanding. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, Berlin, Germany. 2704–2709.

Kok, Kasper. 2016. The status of gesture in cognitive-functional models of grammar. Vrije Universiteit Amsterdam, Netherlands. Utrecht: LOT PhD dissertation.

Krahmer, Emiel & Marc 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, Silva H. 2014. Recurrent gestures. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Jana Bressem (eds.), *Body – language – communication. An international handbook on multimodality in human interaction*, 1558–1575. Berlin: De Gruyter Mouton.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Langacker, Ronald W. 1988. An overview of Cognitive Grammar. In Brygida Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics*, 3–48. Amsterdam: John Benjamins.

Langacker, Ronald W. 2008. *Cognitive Grammar: A basic introduction*. Oxford: Oxford University Press.

Müller, Cornelia. 2004. Forms and uses of the Palm Up Open Hand: A case of a gesture family? In Cornelia Müller & Roland Posner (eds.), *The semantics and pragmatics of everyday gestures: The Berlin conference*, 233–256. Berlin: Weidler Buchverlag.

Müller, Cornelia. 2008. What gestures reveal about the nature of metaphor. In Alan Cienki & Cornelia Müller (eds.), *Metaphor and gesture*, 219–245. Amsterdam: John Benjamins.

Müller, Cornelia, Jana Bressem & Silva H. Ladewig. 2013. Towards a grammar of gestures: A form-based view. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Sedinha Teßendorf (eds.), *Body – language – communication. An international handbook on multimodality in human interaction*, 707–733. Berlin: De Gruyter Mouton.

Müller, Cornelia & Susanne Tag. 2010. The dynamics of metaphor: Foregrounding and activating metaphoricity in conversational interaction. *Cognitive Semiotics* 6. 85–120.

Namy, Laura L., Aimee L. Campbell & Michael Tomasello. 2004. The changing role of iconicity in non-verbal symbol learning: A U-shaped trajectory in the acquisition of arbitrary gestures. *Journal of Cognition and Development* 5(1). 37–57.

Oakley, Todd. 2009. *From attention to meaning*. Bern: Peter Lang.

Rosch, Eleanor & Carolyn B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7. 573–605.

Schmid, Hans-Jörg. 2007. Entrenchment, salience and basic levels. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 117–138. Oxford: Oxford University Press.

Schoonjans, Steven. 2014. Modalpartikeln als multimodale Konstruktionen. Eine korpusbasierte Kookkurrenzanalyse von Modalpartikeln und Gestik im Deutschen. Unpublished PhD dissertation.

Sperber, Dan & Deidre Wilson. 1986. *Relevance: Communication and cognition*. Oxford: Blackwell.

Steen, Francis & Mark Turner. 2013. Multimodal Construction Grammar. In Michael Borkent, Barbara Dancygier & Jennifer Hinnell (eds.), *Language and the creative mind*, 255–274. Stanford, CA: CSLI Publications.

Streeck, Jürgen. 2009. *Gesturecraft: The manu-facture of meaning*. Amsterdam: John Benjamins.

Weigand, Edda. 2010. *Dialogue: The mixed game*. Amsterdam: John Benjamins.

Wittgenstein, Ludwig. 1953. *Philosophical investigations*. Oxford: Basil Blackwell.

Wu, Suwei. In Preparation. Transitivity, multimodality and gesture. PhD dissertation, Vrije Universiteit Amsterdam, Netherlands.

Zima, Elisabeth. 2014. English multimodal motion constructions. A construction grammar perspective. *Papers of the Linguistic Society of Belgium* 8. 14–29. http://uahost.uantwerpen.be/linguist/SBKL/sbkl2013/Zim2013.pdf.