DE GRUYTER

Research Article

Gillian Ramchand*

Verbal Symbols and Demonstrations Across Modalities

Abstract: In this paper, I develop a new theory of the ingredients of semantic composition for the verb phrase, building on work in Ramchand (2018). I argue that the essential properties of this new approach make possible a new rapprochement between the theoretical analyses of symbolic verbal meaning across the visual and auditory modalities. The innovation of the new theory is that it partitions the verb phrase into a lower purely symbolic zone and a higher instantiational situation zone, mediated by the demonstrative act (cf. Davidson 2015, Henderson 2016). In the first part of the paper, I lay out the system, and then I show how it can be used to give a formal analysis of gestural and iconic elements in language. Along the way, I discuss the different ways in which iconicity in the two modalities plays out.

Keywords: iconicity; demonstration; co-speech gesture; quotation; event kinds

1 Motivations for a New Ontology

In this short paper, I introduce a new theory of the ingredients of semantic composition in the verb phrase that I recently proposed in Ramchand (2018), and argue that although it was proposed in that work to deal with a central problem in the cartographic layering of meaning, it turns out to be well suited to creating a common system for integrating symbolic content in human languages across modalities. Specifically, the design features of the solution proposed allow the simultaneous integration of both descriptive and demonstrative content (in the sense of Davidson 2015), and can be adapted to both onomatopæic and co-gestural effects in spoken languages. The system also allows one to formulate explicit hypotheses concerning the commonalities and differences between languages in the spoken and signed modalities respectively.

In this introduction, I first lay out the cartographic and cognitive motivations for the proposal made in Ramchand (2018). The second section presents the technical aspects of the proposal in detail and shows how it solves the problems raised in the introduction. The third section applies the framework to the integration problem for parallel modalities, formulating a technical vocabulary for combining parallel demonstrative and descriptive content, and integrating it with the extensional semantic system. The final section discusses the implications that this kind of system has for the compositional semantics of signed languages and what similarities and differences it predicts with respect to spoken languages.

1.1 Cartographic Generalizations

Cartography tells us that there are robust crosslinguistic generalizations about the ordering of meaning elements in an extended functional projection. Specifically, in the verbal domain we have robust evidence for a universal functional sequence, or extended projection rooted in the lexical verb (1).

*Corresponding Author: Gillian Ramchand: UiT The Arctic University of Norway; Email: gillian.ramchand@uit.no

(1) $\text{Mood}_{speechact}$ > Mood_{evid} > Mod_{epist} > T > Mod_{circ} > Asp > Voice > Cause > V
         Cinque (1999), Grimshaw (1991), Julien (2000)

Even those who are uncomfortable with the universalist claims coupled with fine grained 'cartography', nevertheless subscribe to the C > T > V template of extended verbal projections and language specific rigid ordering (see also Ramchand and Svenonius 2014).

Moreover, at the bottom of every functional sequence, we find evidence for a kind of substantive, conceptual, rich, yet flexible kind of meaning, as denoted by open class items. This is often expressed in theories as a difference between lexical (open class) and functional (closed class) items, and extended projections in the sense of Grimshaw (1990) are all stated with the lexical formatives at the base. Evidence for this kind of *layered meaning* is pervasive and exceptionless crosslinguistically, and yet it currently looks 'accidental' or 'templatic' from the point of view of our formal ontologies. While syntacticians perhaps implicitly believe that a deeper semantic understanding will eventually release them from templatic formulations, it is clear to the formal semanticists that their own ontologies and primitives do not provide a ready made solution for meaning layering. In fact, morphosyntactic layering is seen by semanticists as a *syntactic* fact, something that their formalism can track or follow, but not something that will be explained by that formalism.

The semanticists are right. They use extensional formal ontologies where situations themselves, or referents, are fully specified particulars, and there is no semantic reason why some features of a situational particular should be linguistically privileged over any other. Situations are particulars that exist in worlds and at times. Events are minimal subparts of situations (the kinds expressed by vPs, according to Kratzer 2014), but are full real world particulars, objects in the extensional world. The fact that events are particulars with both causal and temporal properties for example, means that these two pieces of information are not distinguished in a principled way. So the fact that causational information occurs hierarchically lower in the verbal extended projection than temporal information in actual languages is not a property of the situation per se, for those properties are not intrinsically ordered as far as the real world particular is concerned. This makes the internal structuring of propositions as in (2) a templatic matter for the semanticists as well.

Looking more closely at the verbal extended projection we see that over and over in the literature on the interface between syntax and semantics, we find evidence for the existence of *event kinds* which are built up prior to the information about a particular situation in the world. For example, Gehrke (2013), Gehrke (2015) and subsequent work has argued from the interpretation of adjectival passive participles for the existence of event kinds as a conceptual primitive, paralleling the conceptual primacy of kinds in the nominal domain (Carlson 1977, Zamparelli 2000, Chierchia 1998, Hinzen 2016). The domain of argument structure is also a place where generalizations over participant relationships are encoded, as a way of classifying different events. Argument structure and aktionsart more generally, are part of the lowest domain of verbal information and seem to be organized according to fairly robust and universal structuring principles (see Ramchand 2019 for a discussion and summary.) Crucially, all of these general properties of verbal meaning are *independent of the actual referential participants*, or of the *time and location* that the events occur at. This paper will not go into any detail with regard to the specific structuring principles underlying argument structure, or whether they manifest themselves in the same way in both spoken and signed languages. Rather, it will address the issue of the low zone itself, and how it is to be distinguished from the linguistically higher domains that provide layers relating to event instantiation and reference (see also Hinzen 2016).[1]

The formal semanticist does have certain formal tools in place to talk about the abstract concepts of 'kinds' in both the nominal and verbal domains. But in both these cases (nominal kinds and event kinds) they are built by generalizing over particulars, with the aid of the device of possible worlds (Lewis 1986). In fact, with the exception of Gehrke and collaborators, the work cited in these domains all use some kind

---

[1] The conclusion will be in fact that languages with more salient iconic content provide the clue to how and why such meaning layering should be handled, and the shoe will be on the other foot with regard to whether signed languages are like spoken languages in the sense that we will be forced to look at spoken languages through the lens of demonstrated content. We will return briefly to the question of argument structure commonalities across modalities once the overall structure of the proposal in place.

of modal/possible worlds apparatus to formally describe kind level meanings, and systematically represent kinds as being built out of particulars.[2]

If semantics is to follow morphosyntax in such a way that what is simple and underived in the one system corresponds to what is simple and underived in the other, then *essence* must precede *existence* in the cumulative building up of a natural language proposition. This is because the functional sequence tells us that there is some notion of kind/property that resides lower down, close to the root and is the basis for the build up of reference to particulars. So in morphosyntax, particulars are built out of essences, exactly the inverse of the ontological commitments of the extensional semantic model. If we believe that semantic theories constitute a purely symbolic level (in the sense of Marr 1982) for representing the meaning of syntactic structures, and that there is no necessity for the ingredients or ordering of the primes of that symbolic representation to track either the syntactic primes, or the primes of actual cognitive composition in human brains, then there is no need to be disturbed by this fact. If we wish to have a symbolic system that is in a position to straightforwardly begin to generate predictions and hypotheses about the algorithmic level of meaning composition, then are are serious grounds for concern.

1.2 Internalism meets Externalist Semantics

I will argue that the other factor we need to consider is what might be thought of as a 'third factor' property related to the nature of language itself. To see how this plays out, we need to revisit another old debate concerning internalist vs. externalist considerations which is directly responsible for the inverse ontological commitments of the semanticist and the syntactician.

The externalist argues that there is a crucial 'aboutness' to language, and that if we attempt to ground our theories in internalist notions then we are condemned to theories that make no sense of the intersubjectivity of language and which end up being at best unfalsifiable, and mystical at worst. The internalist on the other hand is trying to account for the fact that language is represented in the mind-brain of actual individual speakers. We would like to know what precisely it is that they have memorized, and how it is deployed in a particular linguistic context.

With respect to our internal representations, I think that Chomsky is entirely right when he points out that mental phenomena, and the meanings of our natural language symbols, do not seem to be explicable in purely extensional terms, not even with the help of Lewisian possible worlds. Citing Nagel 1993, Chomsky argues that:

> It is a hopeless task to "complete the materialist world picture" by translating accounts of "mental phenomena" in terms of a "description that is either explicitly physical or uses only terms that can apply to what is entirely physical" or perhaps give "assertability conditions" on "externally observable grounds". (Nagel 1993. pg.37)
> Chomsky 1995. pg 4

Instead, Chomsky advocates a more naturalistic enquiry into the internal products of a natural language system, which will be consistent with the position I will take in what follows.

> Neurologist Rodolfo Llinás (1987) puts the matter well when describes perception as "a dream modulated by sensory input", the mind being a "computational state of the brain generated by the interaction between the external world and an internal set of reference frames" . But the internal frames that shape the dreams are far more intricate and intriguing than often assumed, even at the level of the lexicon, still more so when we turn to expressions formed by the computational

---

2 Event properties, in the work of McNally and Gehrke (Gehrke and McNally 2015, Gehrke 2015, Grimm and McNally 2015), is an attempt to introduce the kind or essence of the event (cf. Fine 2005) as a primitive of the ontology. These authors are very clear that they think of the notion of event essence as *preceding* information about instantiation in the verbal functional sequence. But there are severe problems with formalizing essences directly, in a semantics grounded in particulars. In particular, details of compositionality within the properties domain itself are difficult to make precise. In what follows, I will propose a different strategy, although I am sympathetic to the general research strand, and to the philosophical position of Kit Fine on these matters.

processes.
Chomsky (1995), pg 23

First of all, I will continue to assume a prominent role for events (which I use interchangeably with situations) as real world particulars in the system I envisage, and still maintain the externalist agenda for semantics in the sense of building a system that makes sense of language's relation to the world. Rather than make propositions denote truth values directly, I assume as background a kind of truthmaker semantics for situations (Barwise and Perry 1983, Kratzer 2014, Fine 2014) such that "Something on the side of the world— in this case, a state of affairs— verifies something on the side of language." (from Fine 2014 pg 1). The verbal extended projection builds up a description of a full situational particular, which is then verified, or not, by the current state of affairs.

In order to integrate a truthmaker situation semantics with the internalist agenda, we need to focus on the role of the symbolic primes of a natural language system and what they need to be able to do. Barwise and Perry (1983) point out that, in addition to the 'aboutness' and potential 'intersubjectivity' of language, we need to account for the *reusability* and *perspectival relativity* of the symbolic prime.

2 Utterances as Generalized Demonstrations

In order to make sense of both internalist concerns and meaning layering, we can take our inspiration from an apparently extreme and exotic case in natural language semantics, that of the 'ideophone'. As Henderson (2015) puts it, work on the formal semantics of ideophones is scarce because it is simply not obvious how to integrate descriptive content and depictive content simultaneously in the compositional semantics, and this is precisely what grammatically integrated ideophones seem to do. In giving his own account, Henderson explores a formal foundation for the notion of *demonstrations* from Davidson (2015) and extends it to account for the ideophonic data. Intuitively, demonstrations are a special type of communicative event that stand in a similarity relation with the event demonstrated.

The radical move I will make here is to take the lesson of ideophones as telling us something about the general case. In other words, I will propose that *all* linguistic symbol use should be implemented in terms of a demonstration event, not just obviously ideophonic elements.

The preview of the system I will propose goes like this. For human language to get off the ground, we need to have (i) common possession of symbols that are *abstractions over the different actual situations encountered in the learning phase*, and (ii) a speaker to deploy those symbols as a means of characterizing new situations in the world as she comes across them. Symbols do not denote 'essences" as in some attempts to capture this intuition, rather they are themselves objects in the ontology, artifacts of human language itself. Symbols in turn are *partial* descriptions of events (worldly particulars), but because it is the symbol that is being combined and manipulated in the lowest zone of compositional combination, no unwanted existential implications (in this, or any other possible world) are generated.

In order to do this we need to add to the usual model, a domain D_μ which is the domain of well-formed linguistic entities of type μ, after Potts (2007). These linguistic objects are triples, consisting of a < `phonological string`, **syntactic features**, SEMANTICS>. Full expressions of type μ will be written in `sans serif` . So for example, the verb `run` might have the denotation:

⟦ `run` ⟧ = < `run`, < **init, proc** > , $\lambda e[run(e)]$ > where we represent the content of the third, semantic, position in terms of a (selective) property of events given here for familiarity and convenience in the lambda notation. It is important to state here, that the lambda notation is somewhat misleading here since the third member of the triple is probably more complex and polysemous than this, and 'meanings' of lexical items in this sense are not literally functions. I abstract away from these particular complexities here for the purposes of focusing on the architecture of the proposal.

For convenience, we adopt the convention in Henderson (2015) which uses the bottom corner notation to pick out the semantic part of the triple denoted by something of type μ. Thus, ⌞ `run` ⌟ = $\lambda e[run(e)]$

It is important here that the 'meanings' of the symbols themselves need to be devoid of temporal or worldly information. They form the hierarchically inner core which is then clothed with the contingent information of time, place and world, to link descriptions to actual particulars. The symbols of a person's language are acquired through actual experience of the world, both sensory and cognitive, but they are, importantly, abstractions over the particular instantiations exposed to. This abstraction process is probably partly mechanical but definitely automatic, and certainly filtered through our general cognitive predilections concerning what we consider worthy of attending to. The act of abstracting away from the inessential in a situation to form a representation of the properties that are to be encoded by the symbol being learned is a complicated question way beyond the remit of this paper, and is indeed a central issue in the field of language acquisition. However, one thing is clear—- this process is a natural and immediate one and takes place (perhaps imperfectly) on the basis of one or two exposures. The result (which may be fine tuned over the life of the speaker) is a partial description of essential properties for the event that, on grounds of reusability, are *necessarily* devoid of information related to the particularity of the instantiation (time, place and the reference of the participants). The remarkable thing about possessing an inventory of symbols, which are then consciously deployed in multiple situations, is that this must go hand in hand with the instinct to abstract away from the particular to represent the essential, or reusable content of that sign. It is this third factor property of the symbolic system itself that leads to temporal and spatial, and referential information being represented morphosyntactically external to the memorized sign.

Summary: Lexical Symbols as Elements of the Ontology

(a) Symbols of the language constitute the domain D_μ, which are triples consisting of a `<`phonological string`,`**syntactic features**, SEMANTICS`>`

(b) The semantics of a verbal lexical item are partial descriptions based on sensory and cognitive abstractions over experience.

(c) The syntactic part of the information in a triple that is a member of D_μ, is a subtree of the language. The merge of $u_1 \in D_\mu$ and $u_2 \in D_\mu$, creates a derived element of D_μ, u_3, which has the syntactic representation built by merging the syn-rep of u_1 with the syn-rep of u_2, and a semantics is composed by ordinary argument identification of $\llcorner u_1 \lrcorner$ and $\llcorner u_2 \lrcorner$.

After the symbolic layer (possibly complex) has been constructed, it is necessary to convert these partial descriptions (properties of symbols) into a property of events. This is done by means of an explicit deployment operation. Let us take a look at how Henderson (2015) makes the conversion in his work on ideophones. The conversion makes use of a demonstration event d and a predicate TH_δ which combines with a (possibly complex) symbol to give the 'theme' of that event, namely the use of a particular linguistic object u. The DEMO predicate relates d and a general event variable e, saying that d 'demonstrates' or has certain structural properties in common with e.

(2)   QUOTE : $\lambda u \lambda d \lambda e [TH_\delta(d) = u \wedge DEMO(d,e)]$

In my own proposal, I will carry over this QUOTE function to the general case and propose that the reusable essential symbolic content of any perfectly ordinary sign is the equivalent of Henderson's ideophone. In other words, a symbol is a conventionalized ideophone used to invoke and describe an event. In order to make this fully general, we need to replace Henderson's DEMO predicate with what I will call CONVEY, and replace the exotic idea of a demonstration event with the ordinary garden variety UTTERANCE act itself. I replace the δ subscript on TH with α since we are no longer talking specifically of demonstrations. Thus, Ramchandian 'quotation', built at the level of EvtP (my label for the highest head in the first phase domain, equivalent to VoiceP in other systems) will have the following general form (I), which reduces in the limit to the ascription of partial descriptive content (II).

Deployment of Symbolic Content at EvtP

I. EvtP : $\lambda d \lambda e[\text{UTTERANCE}(d) \wedge \text{TH}_\alpha(d)=u \wedge \text{CONVEY}(d,e)]$
Property of an UTTERANCE event d and event e, which has u as its theme, and where d deploys u ($\in D_\mu$) to convey e.

II. In the case of purely conventional (i.e. non-depictive) LIs, uttered with sincerity and without metaphor or hyperbole,
'$\text{TH}_\alpha(d)=u \wedge \text{CONVEY}(d,e)$' $\longrightarrow$ '$\llcorner u \lrcorner (e)$'

So in other words, at the level of EvtP, which I assume is the label for the edge of the symbolic zone, we get the construction of a property of speech events d and situations in the world e such that someone uses that speech event to convey e. The situation e is conveyed in part by deploying the symbol of the language (here notated with the label u). The second part of the definition says that in the case of purely conventional (non-ideophonic and non-iconic signs), the CONVEY operation by an UTTERANCE simply means that the speaker intends that the semantic partial description connected to the symbol does indeed apply to the event in question. However, because we have explicitly represented the speech event, we leave it open that as a communication event it can do its 'conveying' in other ways as well. We will see that this will become useful as we seek to integrate iconic content into the event description in section 3.

This 'metalinguistic' move is a representational encoding of the intuition that reference involves a speaker and a context in addition to the symbol she is deploying. But it is not just a matter of a speaker X using the symbol Y to refer to the object Z, we need to leave room also for the contextual circumstances and mode of deployment of the symbol in question. Once again Chomsky (1995) puts it well,

> More generally, person X uses expression E with its intrinsic semantic properties to talk about the world from certain intricate perspectives, focusing attention on specific aspects of it, under circumstances C, with the "locality of content" they induce (in Bilgrami's sense).
> Chomsky (1995), p. 43

Before we turn to a more explicit treatment of the interesting implications this system has for integrating iconic content in both signed and spoken languages, there is one further aspect of the system that needs to be put in place.

Recall that the one of the motivations for creating this lower symbolic zone in the VP was the fact of morphosyntactic layering across language, and the existence of symbolic/essential content at the core of all actual linguistic meaning building. Once the symbolic layer has been 'deployed' by the speaker, the complex representation so formed is now ready to be directly modified to receive temporal and spatial information related to the actual instantiation of the event/situation being described. In my own implementation, I follow Champollion (Champollion 2015) in going for a quantificational event semantics, where the event variable itself is existentially bound quite low (at the edge of the VP) and where what is carried up for modification is a variable over *properties* of events. This is independently necessary, as Champollion argues, because the event variable always scopes underneath other quantificational elements. To illustrate, here is Champollion's denotation for the verb phrase *see Mary*.

(3) $[\![$ see Mary $]\!]$ = $\lambda f \exists e[\text{see}(e) \wedge f(e) \wedge \text{th}(e) = \text{Mary}]$

The verb phrase now denotes a property of event properties (here indicated by the predicate variable f), a move that is required to allow further properties of the event to be added after existential closure. The locus of closure of the event variable converges in my own implementation with the introduction of the variable f representing *spatiotemporal/worldly properties* of events which I propose is a complex property relating d and e. Put another way, f is a property of events e, *anchored in* d. Thus, the AspP built up by my quotational quantificational system, will therefore look as in (4).

(4)   ⟦ AspP ⟧ = $\lambda f_{<v,<v.t>>}\lambda d\exists e[\text{UTTERANCE}(d) \wedge \llcorner u \lrcorner(e) \wedge f(d)(e)]$

So at the level of AspP we have a property of relations f that link the UTTERANCE context d with an existing event that is being demonstrated, or described in d. That event has conceptual/perceptual properties as characterized by u. At this point temporal information can be added to the event description that was impossible before. The quantificational aspect of the compositional system is important for the system defended and described in Ramchand (2018) and the data analysed there, we will see later on that it will also be useful in distinguishing between two different types of iconic content in section 3.

Now, semanticists understand very well the need for incorporating contextual information to build meanings that have actual truth conditions, and the relevance of the UTTERANCE itself (the kaplanian context so to speak) is especially crucial for the conversion of indexical elements with a 'character' to what Kaplan calls 'content' which can then be verified, or assigned truth conditions. One way of thinking about the proposal here is that the symbolic zone represents kaplanian 'character' (most obviously different from 'content' in the case of so called indexical elements), but which is now the definition of the whole lower symbolic zone. In fact the system of composition here more closely tracks the order of composition proposed by Kaplan (1989) than the standard formal semantic treatments which treat context as a parameter of the model against which a whole proposition is interpreted.

The semantics that makes crucial reference to UTTERANCE context is not controversial, but reifying the speech event this low down in the structure certainly is. One might wonder why the speech event should be reified in this way at all. Why put it in the representation in this literal fashion, instead of simply invoking it in the model or in the process of interpretation? Relatedly, why represent deployment explicitly and treat language symbols as elements of the ontology in this linguistic representational way?

The full justification for this way of building up the compositional semantics can only be seen in the extent to which it proves a productive and fruitful way of looking at old and new language puzzles and in generating new predictions that can be verified. One such example of productive extension will in fact be explored in section 3. But in addition, as the closing salvo for this section, I offer a few more general and conceptual considerations in favour of the radical rethink I am proposing.

The first conceptual consideration is that by reifying the symbolic zone and making claims about what exactly is being represented there, we get a better understanding of the representation of lexical items, and systematic polysemies and reusabilities that are exploited for lexical symbols. By isolating this area of lexical symbology (in distinction to syntactic features or functional items), we create a more algorithmic theory of how meaning is built up psycholinguistically with claims and predictions that are more commensurable with the primes being investigated by psycholinguists. The underwriting of the natural language generalizations in terms of meaning layering was the primary motivation for the system as proposed in Ramchand (2018) and the fact that it reduces one robust aspect of the V < T < C template to virtual conceptual necessity, given conditions of reusability, was a strong point in its favour.

To summarize, then, in Quantificational Quotational Semantics (QQS) we build up a representation of a situation in four stages, as shown in the box below.

> **QQS: Layered Semantic Composition in the Clause**
> (i) The putting together of lexical items which encode certain event properties. This stage needs to be productive and compositional, but with no reference to temporal or world parameters. (The Interior)
> (ii) The assertion by the speaker of the existence of an event with those properties. (Deployment)
> (iii) Addition of temporal and world properties to the event. (Referential/Instantiational domain)
> (iv) Anchoring of the worldly and temporal properties via the Origo (the speaker and her contextual coordinates).
> (see also Hinzen and Sheehan 2015)

3 Applying QQS to Demonstrative Content

It is perhaps not surprising that this system is well suited to expressing demonstrative (as opposed to just descriptive) content because it was originally designed to integrate the iconic content of ideophones into the descriptive system. Henderson's inspiration in turn came in part from the influential paper by Kathryn Davidson (Davidson 2015) who was explicitly concerned with integrating gestural iconic content within the signed modality. One of the very important aspects of the latter paper is the recognition that demonstrated and descriptive content are deeply integrated in the meaning of natural language symbols. This is nowhere more evident than in the symbols deployed by users of signed languages. Many have argued that sign languages employ iconic effects to a greater degree than spoken languages do, possibly due to the richer expressive power of spatial information when it comes to delineating fine grained properties of situations (Wilbur 2008. Schlenker 2018, Goldin-Meadow and Brentari 2017). As Schlenker (2018) emphasizes, sign language symbols tend to do description and depiction *at the same time*. He formulates the following hypothesis concerning iconicity.

> **Hypothesis 2: Iconicity**
> Sign languages make use of expressions that simultaneously have a logical/grammatical function and an iconic semantics, defined as a semantics in which some geometric properties of signs must be preserved by the interpretation function. Schlenker (2018), pg 129

This fact poses a challenge for formal semantic theories, since up until Davidson (2015) there had been relatively little attempt within this literature to engage with and integrate this sort of content within a compositional semantic system designed for descriptive content. In Davidson (2015) both these aspects of meaning could be fed in to the compositional semantic system in a way that respects the hierarchy of the morphosyntactic vehicles conveying it, and this was very important. Her innovation involved reifying the notion of a demonstration[3], the innovation which is carried over to Henderson's implementation of the ideophonic case. Indeed, the simultaneity of descriptive and depictive content is also true of symbols in spoken languages, where the conventionalized symbol can encode iconic elements in the form of sound symbolism and onomatopœia, even in languages that are not usually claimed to possess 'ideophones' (see Dingemanse et al. 2015 for a recent overview).

In generalizing the system as above, I am claiming that the ideophones are just a special case of what is always true, namely that a symbol is deployed by a speaker in a context to create a description of a real world event. In the limit, the deployment event is just the act of uttering. But what if the UTTERANCE event is fancier than that? What happens if the UTTERANCE event comes with a particular intonation, co/speech gestures like eye rolls or hand movements? What if the nature of the sign deployed carries presuppositions telling us that the deployment event is iconically similar to the event being evoked? We need to ensure that by generalizing the Henderson/Davidson/Potts treatment of ideophones to conventional descriptive items, we still retain the flexibility to account for the more depictive kind of content.

Because of the reification of the deployment event d, we allow ourselves access to a variable on which to express differences of this kind. The CONVEY predicate was chosen to be maximally general, but we *could* state principles which choose different flavours of 'conveying' depending on the nature of the symbol involved and how it is being deployed. Let us take a look again at the general schema for denotations of EvtP in QQS, repeated here in (5).

---

3 Davidson (2015) also uses her system to account for role shift more generally in signed languages, and relates this to quotation in written and spoken languages. I will not be discussing the more explicit version of natural language quotation in this paper, although I do believe it to be a different phenomenon from the internal quotation I am advocating quite generally. Very briefly, true 'external' quotation is *non-integrated* grammatically, while the formal system of inner quotation proposed in QQS crucially is integrated. My view on external quotation is that it involves not context shift, but the transformation of the speaker— via play acting, or literally role shift.

(5) EVTP : $\lambda d\lambda e[\text{UTTERANCE}(d) \wedge \text{TH}_\alpha(d)=u \wedge \text{CONVEY}(d,e)]$
Property of an UTTERANCE event d and event e, which has u as its theme, and where d deploys u ($\in$ D_μ) to convey e.

In the boring case (i) the UTTERANCE is straightforward and sincere and has no significant other properties and (ii) the symbol deployed carries straightforward conventionalized descriptive content. I claimed in Ramchand (2018) that in this case, the formula can be reduced as follows.

(6) $\text{TH}_\alpha(d)=u \wedge \text{CONVEY}(d,e) \equiv$ '⌞ u ⌟ (e)' (pure DESCRIPTION)

One option would simply be to introduce a special relation DESCRIBE in this case to link d and e.

(7) EVTP : $\lambda d\lambda e[\text{UTTERANCE}(d) \wedge \text{TH}_\alpha(d)=u \wedge \text{DESCRIBE}(d,e)]$
Property of an UTTERANCE event d and event e, which has u as its theme, and where d deploys u ($\in$ D_μ) to describe e.

Continuing in this vein, we would have to invent a correspondingly specific predicate for symbols that contributed in a special way by also containing iconic content. This is in fact the intention behind the Hendersonian predicate DEMO. In his implementation, specifically in the case of demonstrations, the symbol in question *carries a crucial presupposition* that the UTTERANCE in addition bears a similarity relation to the event being conveyed (in Henderson 2015 's formulation). In this case, the relation between d and e is more of an explicit demonstration of the properties of e. We might represent this more specifically as in (8)

(8) EVTP : $\lambda d\lambda e[\text{UTTERANCE}(d) \wedge \text{TH}_\alpha(d)=u \wedge \text{DEMO}(d,e)]$
Property of an UTTERANCE event d and event e, which has u as its theme, and where d deploys u ($\in$ D_μ) to demonstrate e, because d bears some kind of similarity relation to e.

But in principle similarity could be at the level of sound (onomatopeia) or at the level of spatial iconicity, as in the case of many sign language symbols. We could therefore further propose more fine grained predicates relating d and e, depending on the nature/modality of the similarity evoked.

(9) DESCRIBE: (the symbol used in) d has a conventional semantic associated meaning that usefully describes e.
DEMO (iconic/auditory): (the symbol used in) d demonstrates or has certain auditory properties in common with e (onomatopeia).
DEMO(iconic/visual): (the symbol used in) d demonstrates or has certain structural properties in common with e (spatial iconicity).

However, I think this move would be a mistake and begins to undermine an intuition that I think is important here, namely that conventionalized symbols of a language whether in the auditory or visual modalities, can often describe and depict *simultaneously*. The problem is that these different cases are not really cleanly separable from each other in practice. Moreover, the degree of iconicity involved seems to be dependent more on the lexical symbol itself than on the speaker, or contextual or real world knowledge. I propose therefore that the best formal representation uses the general predicate CONVEY which is neutral with respect to the division of labour between demonstration and depiction. The use of the general predicate CONVEY emphasizes the commonalities between the deployment of signs across modalities and regardless of degrees of iconicity.

I have claimed that many instances of CONVEY *combine* different ways of representing the event in question. For example, in languages in the auditory modality, onomatopæia is a common phenomenon, but exists alongside descriptive content. Consider for example the case in (10) where we have a verb *whoosh* that expresses an unaccusative motion event without telos, together with a phonological representation that imitiates the sound of the action.

(10) The water whooshed into the room.
EVTP : $\lambda d\lambda e[\text{UTTERANCE}(d) \wedge \text{TH}_\alpha(d)=$ whoosh $\wedge$ '⌞ whoosh ⌟ (e)' $\wedge$ '⌜ whoosh ⌝ has properties in

common with the sound of e']
(Where I use top corner brackets by convention to pick out the phonological part of the triple corresponding to a symbol of D_μ.)

So, the claim here is that degree of iconicity is *not* built into the choice of predicate relating d and e, contra Henderson (2015), but to the presuppositional properties of the symbol deployed, which in turn affects the nature of the CONVEY predicate mediating d and e.

Here though, we come to a second way in which demonstrated, gestural, or otherwise iconic content can be integrated into the event description. As Goldin-Meadow and Brentari (2017) point out, even in sign languages, a principled distinction can be made between conventional iconic properties of a sign, which are more categorical, and more general gestural effects that can accompany a signed utterance and which are more continuous. In fact, iconicity in signed languages is also found systematically outside of symbolic content, for example using loci in signed space for reference tracking of participants and even time intervals (again see Schlenker 2018 for an overview). For this species of gestural content, and for a range of effects in spoken languages as well such as intonation, eye-rolling, and co-speech gestures, we need to employ a different strategy, but one which still requires the deployment event d in the representation. These cases are also explicitly given an account in Davidson (2015), and my own account is very similar to her basic intuition. I differ from her in strictly separating the conventionalized iconicity encoded in a symbol's denotation, which comes as a presupposition from the symbol itself, from the independent properties of the demonstration event itself. It is not always easy to separate these two effects in a sign language which employs the visual modality for both conventionalized and expressive content. In the former case, however, I would predict that the depicted content is integrated obligatorily, automatically and more categorically. In the latter case, the effects are more relative, and continuous, and contextually variable. Goldin-Meadow and Brentari (2017) show convincingly that these different kinds of iconic effect have different cognitive and psycholinguistic consequences. Interestingly, handshape in ASL seems to be interpreted more categorically than either motion or location. So even though both kinds of depictive effects need to be integrated with the formal semantics for descriptive content, I will argue that it should be done in rather different ways.

I give examples of some of these different kinds of scenarios in turn from spoken language, with an indication of how I think the information should be represented at EvtP. Sarcastic intonation on the symbol itself is used to express speaker doubt about the appropriateness of the symbol deployed. In the case of descriptive content this amounts to an expression of speaker disbelief in the applicability of the symbol's descriptive content to the event being described.

(11)   *Dialogue 1: Sarcasm*
        Speaker A: What did you have for dinner yesterday?
        Speaker B: I am trying to lose weight, so I had a huge bowl of kale with some lentils mixed in.
        Speaker A: Ooh. That sounds delicious. (where red text indicates sarcasm alert intonation).
        EVTP : λdλe[UTTERANCE(d) $\wedge$ TH$_\alpha$(d)= sound delicious $\wedge$ SPEAKER-OF-D DOES NOT BELIEVE('⌞ sound delicious ⌟ (e)')]

When it comes to co-speech gesturing, the demonstrative content of the gesture can be integrated low down in this system by introducing a modifier of d, much as in Davidson (2015). For consistency, I have added this in (12) by means of another flavour of the TH predicate over d, which I will call TH$_{gesture}$ and represent the content with a place holder shape symbol △. We can state as a meaning postulate that all co-speech gestures which are explicitly timed by the speaker to coincide with the UTTERANCE and integrated with the event d, are used to add 'thematic' properties of d.[4] This means that a unified, but more complex d is now being used to CONVEY e, over and above the particular symbol being deployed. We do not need to change CONVEY in such cases

---

4 Goldin-Meadow and Brentari (2017) emphasize the co-timing of speech and gesture in integrated speech + gesture utterances, arguing that there is a strong evidence that they form part of a single unified production 'plan'.

because it remains neutral, and we should not change the presuppositions of the symbol being deployed, but because d has acquired extra properties, we get an implication/entailment of similarity between d and the event e being described.

(12)   *Dialogue 2: Co-speech Gesture*
       Speaker A: You won't believe what happened when my pipes burst on me last night!
       Speaker B: OMG what was it like?!?
       Speaker A: The water whooshed into the room. (with wave-like hand gesture simultaneous with pronunciation of verb)
       EVTP : $\lambda d \lambda e[\text{UTTERANCE}(d) \wedge \text{TH}_{gesture}(d)= \triangle \wedge \text{TH}_{\alpha}(d)= \text{whoosh} \wedge \text{'}_{\llcorner} \text{whoosh} _{\lrcorner} (e)\text{'} \wedge \text{'}^{\ulcorner} \text{whoosh} ^{\urcorner}$ has properties in common with the sound of e' $\wedge$ 'The performance of d has properties in common with the spatial trajectory of e']

In the latter case, we can see that the deployment event d can in principle contain information in addition to the actual symbol being deployed. In the next section, I will explicitly discuss the case of language in the visual modality. In doing so, we will continue to speculate about the different ways in which this system can be used to integrate different kinds of iconic content.

4 The Visual Modality

We have seen that quite generally the reifying of the linguistic symbol as a member of the ontology, and the concomitant reifying of the speech event as a locus of deployment and demonstration, provides a frame that is suitable to both auditory and visual modalities. Unlike other systems of semantic composition, we do not have to squeeze iconic content in right at the end of the compositional process. Rather, in this system, iconic content is integrated early, at the level of the lexical composition, or in the terms of Ramchand (2008), already within the first phase. In this, I think both types of languages are the same, and the system does not in and of itself predict any privileged relationship between the visual modality and iconic content. The system I have proposed for compositional semantics is supposed to be neutral between the auditory and visual modalities, and leaves open the relative division of labour between depictive and descriptive content within the symbolic inventory.

On the other hand, the nature of the visual modality itself gives rise to a number of interesting differences between the two systems that I would not want to lose sight of. One is the fact that iconicity built in to the symbol itself in the signed language is often quite hard to distinguish from the iconicity of accompanying gestures. This is strikingly different in spoken languages where co-speech gestures occur in a different modality from the categorical linguistic aspects. More research will be required before we can know what sorts of consequences this has for the system as a whole. In particular, this fact is likely to affect both language change and metaphoricization in the different modalities. In the case of signed languages, it is a difficult question to disentangle how much of the symbol content of the sign is conventionalized, and how much is dependent on details of the demonstration.

Similar to the treatment of ideophones, one would say that meaning of the sign language symbol can encode both descriptive and depictive information. By way of illustration, consider the hypothetical deployment by a speaker of signed language L_s of the sign $\triangle$ in a language-signing event d. We assume that the conventional symbol $\triangle$ has a particular formal semantic descriptive contribution ($_{\llcorner} \triangle _{\lrcorner}$) possibly including some iconic information. For example, Wilbur (2008) has argued that ASL verbal symbols iconically encode certain event structure properties such as telicity, as a matter of convention. Classifier predicates are cases where the handshape seems to encode certain classificatory properties of the entity being moved or located. These are conventionalized elements and they seem to have iconic content. I assume that by using the sign in question, the signer signals that the descriptive content of the sign holds of the event being described, *and* that the iconic properties as conventionally demonstrated in the sign are also satisfied. In other words, it af-

fects the way in which CONVEY predicate chooses properties from the sign to ascribe to e. This is in principle no different from the onomatopæic verb in English, *whoosh*.

(13)　　EvtP : $\lambda d\lambda e[\text{UTTERANCE}(d) \wedge \text{TH}_\alpha(d)= \triangle \wedge \text{TH}_{gesture}(d)= \triangle \,\&\, \text{CONVEY}(d, e)\,]$

Thus, I think there is evidence from event structure distinctions (Wilbur 2008) and argument structure distinctions (Benedicto and Brentari 2004) within signed languages that in addition to the same kind of meaning layering, spoken and signed languages traffic in the same kinds of internal structural distinctions at this level. One intriguing difference seems to be that some of the content that is merely implicit in the descriptive content of spoken languages, can be expressed as iconic spatial content in signed languages. However, as we have seen, this is not a deep difference of architecture, but an epiphenomenon of the specifics of the two modalities. In fact, what it illustrates (and here I am in broad agreement with Schlenker 2018) is that sign languages often offer clear and overt evidence for universal aspects of linguistic meaning that are harder to detect in spoken languages.

But what about the other gestural properties such as the speed and extent of the sign as demonstrated in that particular context? What effect do these have on the meaning of what is said? The natural thought here is that these additional iconic gestural effects can be integrated as Davidson (2015) suggests in the form of properties of the demonstration event itself directly either as a separate modifying predicate or a special flavour of the TH predicate. As mentioned earlier, I think there are good reasons to sharply distinguish these effects from effects that are conventionalized as part of the symbolic content, even though because of the shared modality, these are often difficult to distinguish in practice.

Although the problem is sharper for signed languages, spoken languages do have their equivalent of the sign vs. gesture analytic question in the case of stress and tone. Intonation and stress can be thought of as co-speech vocal gesturing that creates expressive iconic content. Evidence in favour of this comes from the regularity with which these very devices are used with the same sorts of effects in typologically unrelated spoken languages (one of the diagnostics Goldin-Meadow and Brentari 2017 offer as a distinguishing factor between gesture and sign). Nevertheless, linguistic descriptions of spoken languages regularly involve descriptions of certain lexical items as being obligatorily stressless, or of distinguishing between lexical items that come in stressed or unstressed versions, or items that bear lexically assigned stresses or tones. Also, individual languages can have very specific intonational tunes corresponding to phrasal level meanings, that are highly conventionalized. So here too, there is a blurring between a device generally at the disposal of the speaker to create iconic and expressive content, sometimes bundled into the conventionalized aspects of the symbolic system itself. This is arguably no different from what we see with iconic gestural content encoded in sign language lexical items, side by side with more general gestural effects such as the speeding up of motion, or the distance traveled when signing a particular motion verb.

In fact, I suspect that the co-speech gesturing that is found in spoken languages is very similar to the continuous iconic gesturing in signed languages, but probably overlaps with some forms of categorical symbol iconic content as well. As Goldin-Meadow and Brentari (2017) put it, "sign should not be compared with speech alone, but with speech plus gesture" pg 2.

There is one more feature of continuous or gestural iconic content that is worth pointing out here, and that is the fact that this content is related to the *actual instantiation* of the event in context, and is not given by the deep token-independent properties of the sign. This means, in the system of Ramchand (2018), that this iconic information is contributed at a stage of composition *after* the first phase is complete. To repeat, at the edge of the first phase, the representation of the meaning built to that point is as in (14).

(14)　　$[\![$ AspP $]\!] = \lambda f_{<v,<v.t>>}\lambda d\exists e[\text{UTTERANCE}(d) \wedge \llcorner u \lrcorner(e) \wedge f(d)(e)\,]$

As we can see, in this system d also shows up in the spatiotemporal domain of the clause where spatiotemporal properties of the event f are delineated. I would argue that gestural content that expresses facts about the specific way the event is instantiated in space and time is technically a species of f in the above system (which is a property of relations between d and e) rather than a property of d alone. In other words, the iconic property in question is of the form where f = 'e has spatiotemporal properties similar to the spatiotem-

poral form of d' . In this respect my implementation would in fact deviate somewhat from the one in Davidson (2015), because it is integrated in a rather different larger system, but the dependence on the reification of the demonstration event d remains.

Similarly, I would argue that using loci for tracking individual and temporal referents is a part of the system that delivers properties of f, rather than d alone. The CONVEY predicate relates symbolic content to the event e being described via d, but spatiotemporal properties of e are encoded by f in this system (a spatiotemporal relation between d and e) and it is the second, inflectional phase of the clause where reference and time and properties relating to instantiation are expressed.[5] This means that signed languages use the signing space to iconically represent features of the functional and referential system of language as well as the symbolic and classificatory systems of language. One could speculate that the difference between more categorical iconic effects and more open-ended and continuous ones is correlated with the different hierarchical zones in which the iconicity is generated. I leave investigation of these issues for further research.

5 Conclusion

To conclude, I have argued that symbolic self consciousness and the *reusability* of open class items is necessary for a generative meaning engine. Reusability and sensitivity to speaker perspective are the crucial factors that motivated the quotational system in the first place, and implementing those ideas allowed us to make a start at explaining certain core cartographic generalizations. I am fully aware that this system is a radical departure from previous formal semantic systems and that it requires a readjustment not only in what we consider to be the 'meanings' of lexical items, but also in the traditional divisions between semantics and contextual aspects of meaning. However, the facts seem to warrant such a rethink.

The change in ontology here is designed to reflect the primes actually employed in natural language composition to build meaning, and the topological transformation of the composition system is designed to reflect a more algorithmic view of how human minds compose meaning. To summarize the main features of QQS again, I have argued for:
(i) The reification of the symbol as object of the ontology.
(ii) The reification of the relationship between the deployment event (a communication event situated in the here and now of the real world), early in the compositional history of the build up of the proposition.
(iii) The restriction of the symbol's lexical content to cognitive abstractions over time and locational instantiation.

When it comes to the last point in (iii), the lexical descriptive content of a conventional sign is argued to be modality independent, and hence applies to both signed and spoken languages. However, because of the reification of d, the deployment event, we automatically build in a difference between the signed and the spoken modalities because d cannot avoid having spatial properties in the former case, while it cannot avoid having auditory properties in the latter case. Because d, the utterance event, is not exhausted by the deployment of the conventional content of the symbol, room is also made for demonstrational content to be contributed by d.

We saw that envisaging different presuppositions for symbolic entities could allow us to accommodate both demonstration and description within the same generalized CONVEY relation. At the same time, allowing spatiotemporal properties of events to be iconically demonstrated as well, by providing time locked gestural content in parallel with the symbol deployment, allows both conventional *and* performance related content to be iconic. Neither of these possibilities is tied to a particular modality, so that in principle mixed effects in parallel across modalities (eg. co-speech gesture) are also accommodated.

---

[5] Recall that in Ramchand (2018), this was the underlying reason for building the system in the way it is built—- to capture the fact that causation and argument structure are always encoded hierarchically lower than tense.

No working linguist can fail to be impressed by the important and intriguing differences between the visual and auditory modalities when it comes to natural language. In this paper, I have been arguing for seeing the commonalities rather than the differences, but the framework is also intended to set the stage for more detailed work on each type of language that will allow them to be compared in a more direct and commensurable fashion. The system proposed here has taken the innovations forced on us by the existence of integrated iconic content, and turned it into the new normal. The case of spoken languages with purely artbitrary conventional symbols devoid of iconic content is now seen as a limiting case of a more general compositional system, and not the example that should itself determine the blueprint for natural language meaning composition.

References

Jon Barwise and John Perry. *Situations and attitudes*. MIT press, Cambridge, Ma., 1983.

Elena Benedicto and Diane K. Brentari. Where did all the arguments go? argument-changing properties of classifiers in asl. *Natural Language & Linguistic Theory*, 22:743–810, 2004.

Gregory Norman Carlson. *Reference to Kinds in English*. PhD thesis, University of Massachusetts at Amherst, 1977.

Lucas Champollion. The interaction of compositional semantics and event semantics. *Linguistics and Philosophy*, 38:31–66, 2015.

Gennaro Chierchia. Reference to kinds across languages. *Natural Language Semantics*, 6:339–405, 1998.

Noam Chomsky. Language and nature. *Mind*, 104(413):1–61, 1995.

Guglielmo Cinque. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford University Press, New York, 1999.

Kathryn Davidson. Quotation, demonstration and iconicity. *Linguistics and Philosophy*, 38:477–520, 2015.

Mark Dingemanse, Damian E. Blasi, Gary Lupyan, Morten H. Christiansen, and Padraig Monaghan. Arbitrariness, iconicity and systematicity in language. *Trends in Cognitive Sciences*, 19(1):603–615, 2015.

Kit Fine. *Modality and Tense: Philosophical Papers*. Oxford University Press, 2005.

Kit Fine. Truthmaker semantics for intuitionist logic. *Journal of Philosophical Logic*, 43(2):549–577, 2014.

Berit Gehrke. Still puzzled by adjectival passives. In Cristina Sevdali Raffaella Folli and Robert Truswell, editors, *Syntax and its Limits*, pages 175–191. Oxford University Press, 2013.

Berit Gehrke. Adjectival participles, event kind modification: the case of frequency adjectives. *Natural Language and Linguistic Theory*, 33:897–938, 2015.

Berit Gehrke and Louise McNally. Distributional modification: the case of frequency adjectives. *Language*, 91(4):837–868, 2015.

Susan Goldin-Meadow and Diane K. Brentari. Gesture, sign and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, (doi:10.1017/S0140525X15001247,):1–60, 2017.

Scott Grimm and Louise McNally. The *ing* dynasty: Rebuilding the semantics of nominalizations. In Mary Moroney Sarah D'Antonio and Carol Rose Little, editors, *Proceeding of SALT XXV*, pages 82–102, 2015.

Jane Grimshaw. *Argument Structure*. Linguistic Inquiry Monographs; 18. MIT Press, Cambridge, 1990.

Jane Grimshaw. Extended projections, 1991.

Robert Henderson. Pluractional demonstrations. paper presented at SALT 26, 2015.

Wolfram Hinzen. On the grammar of referential dependence. *Studies in Logic, Grammar and Rhetoric*, 46(1):11–33, 2016.

Wolfram Hinzen and Michelle Sheehan. *The Philosophy of Universal Grammar*. Oxford University Press, 2015.

Marit Julien. *Syntactic Heads and Word Formation: A Study of Verbal Inflection*. PhD thesis, University of Tromsø, 2000.

David Kaplan. Demonstratives. an essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals (ms. 1977, reprinted). In J. Perry J. Almog and H. Wettstein, editors, *Themes from Kaplan*, pages 481–563. Oxford University Press, Oxford, 1989.

Angelika Kratzer. Situations in natural language semantics. In Edward N. Zalta, editor, *Stanford Encyclopedia of Philosophy*, volume 2014 Edition. 2014.

David K. Lewis. *On the plurality of worlds*. Blackwell, Oxford, 1986.

David Marr. *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Company, San Francisco, 1982.

Christopher Potts. The dimensions of quotation. In Chris Barker and Pauline Jacobson, editors, *Direct Compositionality*, pages 405–431. Oxford University Press, 2007.

Gillian Ramchand. *Verb Meaning and the Lexicon*. Cambridge University Press, 2008.

Gillian Ramchand. *Situations and Syntactic Structures: Rethinking Auxiliaries and Order in English*. Linguistic Inquiry Monograph 77. MIT Press, 2018.

Gillian Ramchand. Events and verbal decomposition. In Rob Truswell, editor, *The Oxford Handbook of Events*. Oxford University Press, 2019.

Gillian Ramchand and Peter Svenonius. Deriving the functional hierarchy. *Journal of Language Sciences*, 2014. *Language Sciences*, http://dx.doi.org/10.1016/j.langsci.2014.06.013.

Philippe Schlenker. Visible meaning: Sign language and the foundations of semantics. *Theoretical Linguistics*, 44(3/4):123–208, 2018.

Ronnie Wilbur. Complex predicates involving events, time and aspect: Is this whysign languages look so similar? In Josep Quer, editor, *Signs of the Time.*, pages 217–250. Signum, Hamburg, 2008.

Roberto Zamparelli. *Layers in the Determiner Phrase*. Garland, New York, 2000.