Davidson, Kathryn. “Semiotic distinctions in compositional semantics”. To appear in the Proceedings of the 58th Meeting of the Chicago Linguistic Society, edited by Sam Gray, Quain, Lucas Fagen, Stephanie Reyes, and Irene Tang, 2023

Semiotic distinctions in compositional semantics

Kathryn Davidson
Harvard University

1 Introduction

Within traditional areas of focus in linguistics, the long discussed arbitrariness of the linguistic sign (Saussure 1916) highlights the independence of many linguistic forms from their meaning, such as the fact that the form of the English word *cat* has no particular relationship to any concept of a cat. Put another way, the relationship between the word *cat* and the concept of a cat is entirely by convention, thus arbitary. Arbitariness is often contrasted with iconicity, in which forms bear some relation to their meanings. Examples of iconic linguistic forms come in all modalities including both spoken languages and sign languages, such as the American Sign Language (ASL) word 'cat'[1] which follows the shape of a cat's whiskers, English onomatopeia like *meow* that sound somewhat like the sounds a cat makes, or an English speaker's or ASL signer's gesture of patting a cat that co-occurs with either speaking or signing, all of which involve an expressive form used as part of language, having some non-arbitrary relationship to its meaning.

Although there is recent interest in the extent of iconicity/non-arbitrariness among words across the world's spoken languages (Blasi *et al.* 2016, Winter & Perlman 2021) and a surge of research on both iconic and noniconic aspects of sign languages and gesture (see Goldin-Meadow & Brentari 2017 and replies for overview), there is largely a lack of consideration for the role that iconicity plays in *composing* sentence meaning and interacting with the logical structure of language. For example, if a form A is iconic, and composes with form B, is the iconicity in the resemblance of a form to its meaning interpreted as part of the composite expression ($A + B$), or is it ignored in the interpretation of the composite? How is it affected by logical/functional operators in language like negation (*not*), disjunction (*or*), questions (*does...?*) etc?

Within cognitive linguistic and psycholinguistic approaches, it is often assumed that iconicity is a mechanism for adding meaning/ content, for example via conceptual structure-mapping (Emmorey 2014). Work on iconicity in cognitive linguistics includes many rich observations on how iconicity can both support (Taub 2001) and constrain (Meir 2010) non-compositional meaning like metaphors, in general placing more focus on meaning via iconicity than on meaning via function composition. In contrast, formal semantic approaches foreground semantic compositionality, but mostly ignore iconicity. One line of formal semantic research that does engage with

[1] This image and all other images of single lexical signs from American Sign Language are from the ASL Sign Bank, (Hochgesang *et al.* 2020)

the question of iconic content and compositionality focuses on co-speech gestures, and finds that they differ in their compositional properties from (arbitrary) spoken words, such as their interaction with spoken language negation (Ebert & Ebert 2016, Tieu *et al.* 2017). Formal explanations have generally focused on the issue of gesture and speech existing in separate modalities (Schlenker *et al.* 2013, Schlenker 2018a, Tieu *et al.* 2019), attributing differences in composition to the mis-match between the spoken and visual language modality for co-speech gestures and not on the fact that the gestures are iconic. Esipova (2019), in contrast, argues that the distinction stems from structural considerations, not modality ones, and in fact is quite compatible with the approach taken here, but also does not take iconicity to be a driving factor in these different structural arrangements. Despite differences, all of these formal semantic approaches to gesture in spoken languages work under the assumption that the iconicity of the gestural content has the same *potential* for compositionality as non-iconic content, and in fact treat iconic meaning in gestures within the same kind of formal system used for non-iconic words in the language (Schlenker 2021). Outside of gesture, in related formal semantic work on sign languages the question of iconicity and compositionality has arisen in such varied areas as verb forms (Strickland *et al.* 2015, Kuhn & Aristodemo 2017), adjectives (Aristodemo & Geraci 2018), and anaphora (Kuhn 2020), and the approach has generally been to take iconicity as an advantageous property of the visual modality that allows us to see compositional structure that is present but might be hidden in spoken languages. The assumption again is that the structure in (more iconic) sign and (less iconic) speech is generally the same. Thus we see that in many different approaches to iconicity and sentence meaning, iconic content is often assumed to have similar compositional properties as non-iconic content, often without questioning the role that iconicity itself could play in constraining composition.

The goal of this short paper is to provide some insights into the question of iconic compositionality from a semiotic perspective by dividing iconic content by what it requires to be interpreted, directly following work on depiction in semiotics and psychology (Clark 2016, Dingemanse *et al.* 2015, Dingemanse & Akita 2017, Ferrara & Hodge 2018, Kita 1997), consistent with a wide spectrum of philosophical approaches to depiction that separate its interpretation from the kind of interpretation required by symbolic compositional language (Fodor 2007, Burge 2018). Following in the footsteps of these psycho-semiotic foundations, we will divide iconic language into two categories. On the one hand, there are iconic form-meaning pairings which are symbols that derive meaning via convention: these include onomatopeia like *chirp* in English, expressive sensory ideophones like *gelegele* 'shiny' in Siwu (Dingemanse 2019) or *gorogoro* 'heavy rolling object' in Japanese (Kita 1997), emblems that build on metaphors like the thumbs up gesture for 'good', and visually iconic signs like 'cat' in ASL. We can categorize all of these forms as having **lexical/descriptive iconicity**: the forms are not entirely arbitrary, since we can identify how aspects of the form relate to their meaning, but their meaning is not derived from the form but rather from this conventionalized form-meaning pairing (the stored lexicon), and so it has the *potential* to be represented by an entirely arbitrary symbol. For example, the ASL word and

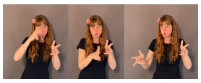
Iconicity in the lexicon	**Depictive iconicity**
(Some) lexical signs: INFORM in ASL	(Some) size gestures: *This huge* *tail*
Classifier handshapes in ASL:	Classifier movements, locations in ASL:
'vehicle'	 'vehicles like this'
Ideophones in Siwu:	Siwu ideophone w/depiction:
gelegele 'shiny'	*gelegele-gelegelegelegelegele*
Onomatopeia:	"Expressive" onomatopeia:
Margola chirped in her cage.	*Margola went "chiiirrrrp!"*
Categorical gestures: 👍	Quotations: *Alexis was like, "No, David!"*

Table 1: Lexical/descriptive iconicity vs. depictive iconicity

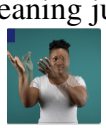
the English word for 'cat' each pick out the same concept; this mapping has to be stored conventionally, and could have been mapped differently, given the different words for 'cat' across both sign and spoken languages of the world, and their sometimes arbitrary nature (e.g. English *cat*). Similarly, *gelegele* and *shiny* might be iconic in having some non-arbitrary aspects of their form (perhaps they seem related to light, for example), but as symbols they contribute a basic meaning via a conventionalized form-meaning mapping.

Iconicity within the symbolic lexicon contrasts with **depictive iconicity** that need not express stable or abstract concepts and thus could not be expressed equally as well by an arbitrary symbol, and instead must be interpreted not via the lexicon (form-meaning conventionalized mappings) but instead via a non-symbolic perceptual mapping of the same sort used to interpret images. Consider, for example, that a painting or a photograph of a cat contributes its meaning through resemblance of looking like a cat. Note that an image is necessarily of a particular cat (in a particular pose, etc) whereas a symbol need not commit to any particular instantiation of the cat kind or concept. In short, depiction involves showing, not telling. As Hodge & Ferrara (2022) note, depiction can accompany descriptive language in both spoken and signed languages, thus it seems possible to talk about "depictive language" as depictions occuring while using language, which can be more and less integrated into a linguistic utterance. Depictive iconicity includes modifications of onomoatopeia and ideophones to mimic a particular sound or perceptual experience, gestures that depict/show a particular size, depictive verbs in sign languages which illustrate a particular arrangement, and even quotation-like demonstrations that depict another's actions, attitudes, or speech (Clark & Gerrig 1990, Davidson 2015), all of which are understood as providing meaning by showing. The interlocutor understands depictions through their own perceptual experiences of what a particular object or event might have looked like, felt like, sounded like, etc.

Descriptive/lexical iconicity involves conventionalized forms mapping to discrete meanings (or an area/network of meetings, when we consider polysemy) in some kind of stored lexicon; depictive iconicity involves neither, and instead meaning derives from mapping a form directly to a particular pictoral/sensorial representation. In many areas of lingustics these have both been categorized as "iconicity", and this conflation seems to have been a source of confusion in both cognitive and

formal approaches. Here we take this semiotic distinction as potentially crucial to understanding the role of iconicity in compositionality. Table 1 provides several contrastive examples: the first column involves forms that bear some relation to their meaning, i.e. they are not entirely arbitrary, but also must be stored in a lexicon and are interpreted as abstract symbols. In the right column are cases of iconic depiction in language, which are not interpreted merely as symbolic but necessarily involve understanding depictions. Notice that there are many cases where (descriptively iconic) symbols from the left column (e.g. classifier handshapes in sign languages, ideophones or onomatopeia in spoken languages) are especially natural starting bases for further using language depictively; see discussion of this point especially in Dingemanse 2015.

2 Depiction and compositionality

We can build upon this distinction between lexical/descriptive iconicity and depictive iconicity to more carefully pose our original question: how should we think about the contributions of iconic language from the perspective of compositional semantics? For **lexical/descriptive iconicity**, the answer appears quite obvious: these are symbols that contribute their meaning just like non-iconic words do. For example, the ASL words 'cat', or  'inform' (which is iconic, in that it shows information coming out of the head) and the English words *cat* or *inform* (which do not have much obvious iconicity) are all symbols. The symbols and *cat* have the same compositional contribution: $\lambda x.\textbf{cat}(x)$ (a function that returns true for entities that are cats).; similarly, the symbols *inform* and have the same compositional contribution: $\lambda v.\textbf{inform}(v)$ (a function that returns true for informing events). Although some are motivated in their form, all of these signs are ultimately abstract symbols, and compose just like other abstract symbols. We see the same thing in onomatopeias like *chirp* or *knock*, which similarly are also abstract symbols (i.e. *chirp* contributes the function that returns true for chirping events, $\lambda v.\textbf{chirp}(v)$), even if they are (descriptively/lexically) iconic in the sense that they are motivated in form and can be quite easily used to support depictions.

While the compositional contributions of lexically/descriptively iconic words are straighforward, it is a wide open question how to think about **depictive iconicity** with regard to compositional semantics. On the one hand, why and how would a picture compose with language? If we were looking at a children's storybook that contained text and pictures, we wouldn't typically view the job of the formal semanticist (or syntactician, or any linguist) to understand how the pictures compose with the text. On the other hand, there has been a recent flourishing of interesting work on precisely this, how to understand pictures alone or pictures accompanying text, by cognitive scientists using the tools of formal semantics that were designed for

natural language. For example, Abusch (2012) and Abusch & Rooth (2022) apply a formal approach to the analysis of narrative picture sequences, Cohn (2012) similarly considers the semantic structure of comic sequences, and Greenberg (2021) applies the notion of entailment that we see in formal semantics to a broad range of pictures. These seem to counter the seemingly default view that pictures fall outside the domain of formal linguistic analysis; applying formal semantic tools to them can be thought of as part of a larger investigation into applying the tools of semantics to domains beyond language (Schlenker 2018b). Thus, how we handle depictive iconicity becomes an important question: do iconic depictions compose (or not) with symbolic structures in language, and if so, how?

As we noted above, much work that investigates compositionality and iconicity takes place within gesture or sign languages, categories defined by their modality, not semiotic distinctions: iconic gestures can involve lexical iconicity and depictive iconicity, and similarly, sign languages involve both lexical iconicity and depictive iconicity. The goal of the current work is to gain some additional insight, then, beyond previous studies looking at linguistic iconicity and composition, by carefully focusing on depictive iconicity, where the stakes are most interesting. Section 3 presents two short quantitative studies investigating the composition of depictive language in written English, a language and modality not typically associated with iconicity. Section 4 presents a proposal based on this data and existing data in the literature, arguing that depictions cannot participate in the construction of alternatives, necessary for composition in many functional areas of the grammar like questions, negations, disjunction, etc. Section 5 applies the proposal to depictions in other domains including Japanese ideophones and depictive classifiers in American Sign Language, as well as English onomatopeia. Section 6 concludes.

3 Naturalness ratings

3.1 Methodology and design

This section describes two short experiments in English to test the hypothesis that iconic language, especially iconic depictions, do not compose with logical operators in the same way as symbolic language. The goal is to understand compositional reflexes of this semiotic distinction. We will focus on the behavior of symbolic language vs. depictions under negation, since negation is in some sense the simplest logical operator, and every human language in the world seems to have a symbolic means of expressing negation. Negation has also been a domain in which compositional differences have been reported between (sometimes depictive) co-speech gestures and words with seemingly similar content in spoken languages. The goal will be to compare depictions and symbols in the context of positive and negative versions of the same logical structure, for example, contrasting the way that depictive vs. symbolic languge composes with positive adverbials (e.g. *always*) vs. adverbials that express negation (e.g. *never*), with the hypothesis that depictions may be less natural than symbolic content in a negative context.

The methodological approach taken here in some sense "a little bit experimental" (Davidson 2020). Instead of hiding the contrast of interest among many fillers, we highlight the contrast by presenting participants with only ten total target sen-

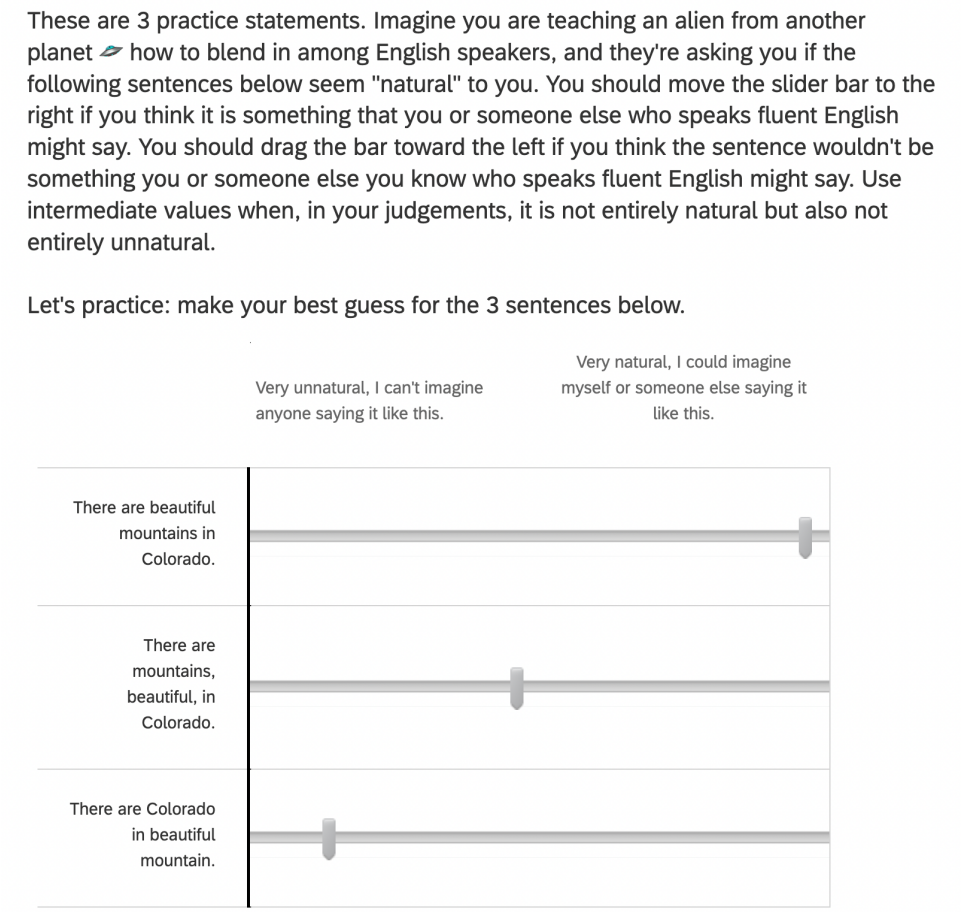


Figure 1: Screenshot of training trials

tences simultaneously, an effective tactic in in previous work in experimental semantics (Marty *et al.* 2020). Participants, recruited via Prolific Academic, were asked to help teach an "alien from another planet" to blend in among speakers in the participant's community. For the alien's sake, participants were asked to rate sentences on a slider bar from "Very unnatural, I can't imagine anyone saying it like this." to "Very natural, I could imagine myself or someone else saying it like this." The goal with this phrasing was to broaden the typical notion of acceptability, which might be difficulty to judge while incorporating depictions, to a task of "naturalness", motivated by the kind of language teaching tasks often used to elicit acceptability judgments by children in language acquisition research. Each mini study involved training with three example sentences (the same for each study): one was natural, one intentionally unnatural (notably: ungrammatical), and one somewhat in between, as shown in the screenshot in Figure (1). Following the training trials, the ten target trials in each mini study were presented in a randomized order. Participants were reimbursed $1.50, an average of $27/hr as estimated by Prolific Academic, the site used to recruit all participants.

3.2 Depiction via vowel lengthening

In the first mini-study, participants (n=52) were presented with a depictive contrast in vowel lengthing. The English adjective *huge* is a (mostly) arbitrary symbol, with little relation between the form and its meaning beyond perhaps some descriptive/lexical iconicity in the choice of low vowels for large things. It contributes its meaning by convention. However, vowel lengthening can be applied to the low vowel to metaphorically convey a larger size using a kind of depictive iconicity, wherein the length of the vowel depicts in a sense the size of a tree. Participants were asked to provide naturalness ratings of 10 total sentences. These were based on five sentence frames, in pairs that varied in whether the adjective involved a lengthened vowel, for a total of ten sentences, which were all presented on screen together in randomized (by participant) order. The sentence frames included a positive episodic sentence, one with a universal quantifier, a negative quantifier, a negative modal, and one intentionally constructed to be not well formed in the language (1).

(1) a. The last time I felt so relaxed, I was sitting under a {huuuge, huge} palm tree in California, enjoying the moment.
(positive episodic)

b. My friend likes the shade so he always sits under a {huuuge, huge} palm tree next to the boardwalk.
(universal quantifier: always)

c. My friend likes the sunshine so he never sits under a {huuuge, huge} palm tree next to the boardwalk.
(negative universal quantifier: never)

d. There are small ones indoors, but I can't find any {huuuge, huge} palm trees in Mass. I don't think they exist.
(negative modal: no exists)

e. I love sitting under {huuuge, huge} in California, the shade feels great.
(ungrammatical)

The hypothesis was, first, that depiction would be most natural in a positive episodic sentence, since the sentence describes a particular event and one could, in part, metaphorically depict aspects of the huge tree through vowel lengthening. Depiction was hypothesized to be worst in negative contexts (like the negative universal quantifier or negative modal) since these involve no entailment that there is an event or object to depict. Thus, depiction should interact with sentence frame, such that sentences without depiction may vary (for example, sentences both with and without depiction are expected to be rated unnatural in the ungrammatical sentence frame), but depiction will be more unnatural under negation than non-depiction sentences, which should be natural in all but the ungrammatical sentence frame.

Data were modeled in R as *glm* with the acceptability (response) as dependent variable and depiction (huge vs. huuuge) and sentence type/environment as independent variables (*Response* $\sim$ *Depiction* * *SentenceType*), with the resulting model plotted using ggplot and shown in Figure 2, and means given in Table 2. Results upheld the two main hypotheses involving depiction. First, depiction was rated highest in the case of the positive episodic ("particular") sentence, in that

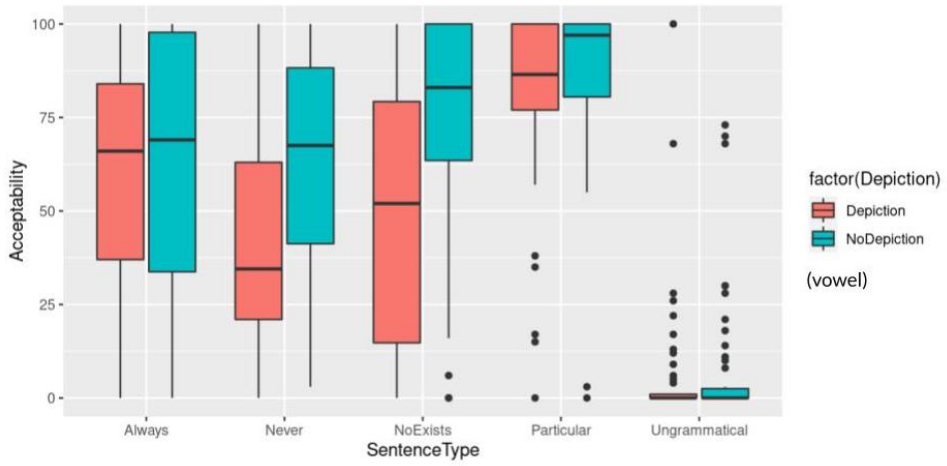


Figure 2: Acceptability ratings of vowel lengthening depictions across five linguistic contexts, based on 52 participants recruited from Prolific Academic platform

Sentence type	Huge	Huuuge
Particular	86.36	80.8
Always	62.6	58.98
No Exist	75.8	49.64
Never	64.58	38.84
Ungrammatical	7.7	6.26

Table 2: Means for Experiment 1

case not significantly different from the sentence without depiction. In the ungrammatical sentence frame, both were rated unnatural, and did not differ from each other. However, depiction diverged from non-depiction sentences in the negative sentence frames, where sentences without depiction were acceptable in the context of negation but sentences with depiction were rated significantly less natural. Note in particular that the positive universal quantifier and negative universal quantifier have exactly the same sentence structure and level of complexity, yet there was a significant interaction between negation and depiction ($p < 0.01$) such that negation decreased the acceptability ratings of (only) sentences involving depiction.

3.3 Depiction via demonstration

The second mini-study also involves depictive iconicity using written language. In this experiment, participants were presented with a depictive contrast in quotation versus an embedded clauses; the first involves depictive iconicity via a demonstration (showing a particular saying/attitude event) while the other involves telling about the same kind of event. The structure followed the study in section 3.2: ten sentences were based on five sentence frames, again involving an episodic particular, a universal adverbial quantifier (always), a negative adverbial quantifier (never), a non-quantificational negative sentence, and an ungrammatical sentence. Both the depictive and non-depictive versions report the speaker's request, but one does so via quotation, the other via an embedded clause.

Sentence type	Demonstration	Embedded question
Yesterday	80.02	95.8
Always	73.51	91.86
Didn't	43.96	78.86
Never	55.64	85.88
Ungrammatical	11.14	16.62

Table 3: Means for Experiment 2

(2) a. At the cafe yesterday, I {was like "could you open the windows?" /asked if they could open the windows.}
(positive episodic/yesterday)

b. At the cafe, I {am always like "could you open the windows?" / always ask if they can open the windows.}
(universal quantifier: always)

c. At the cafe, I {am never like "could you open the windows?" / never ask if they can open the windows.}
(negative universal quantifier: never)

d. At the cafe, I {didn't go like "Can you open the windows?" /didn't ask if they can open the windows.}
(negative/didn't)

e. At the cafe, I them ask {like "could you open the windows?" /if they can open the windows.}
(ungrammatical)

Although the form of depiction was entirely different than the case of vowel lengthening, the predictions were the same: depictions should be most natural in positive sentences, especially episodic sentences, and diverge the most from non-depictive sentences under negation. Both depictive and non-depictive options should be unnatural in the ungrammatical sentence frame.

The hypotheses were again mostly born out in the results, which were analyzed in parallel with the previous mini study. The same *glm* model (here, depiction being quote vs. embedding) plotted in ggplot is shown in Figure 3. The ungrammatical sentence frame was rated as highly unnatural both for depiction and non-depiction versions. Unlike vowel lengthening, there was an overall main effect of depiction, such that quotation/demonstration versions were generally rated as less natural than embedded clause versions. However, the key hypothesis, that there would be an interaction between depiction and negation, was upheld again: depiction diverged the most from non-depiction in the *never* and *didn't* sentence frames. Focusing just on the adverbial quantifiers *always* and *never*, which have the same sentence structure (and thus provide the tightest control), we find a significant interaction between depiction and always vs. never ($p < 0.05$). Means are provided in Table 3. Full anonymized data for both studies are available on the project's OSF website: https://osf.io/y4v9r/?view_only=6e2216967bb14c93bd0fd6f84553b6c2

Overall, it seems that iconic depiction shows a clear compositional sensitivity to negation. This has been reported for gestures (Ebert & Ebert 2016), but we show here that this is less about the modality, and instead about the depiction involved in these iconic forms. In the next section, we discuss why this might be so.

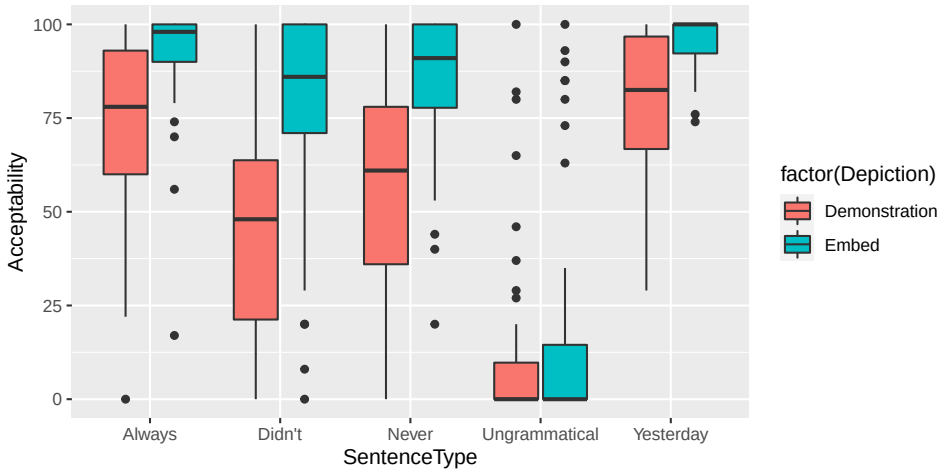


Figure 3: Acceptability ratings of quotational depictions across five linguistic contexts, based on 50 participants recruited from Prolific Academic

4 Interim conclusions: The role of alternatives

Although depictive iconicity is not typically discussed for written language (it is much more obvious in the case of depictive gestures or even spoken language vocal gestures, etc.), it does exist, as in the cases of vowel lengthening and quotation. Our interim conclusion from these two short quantitative studies is that even in written language, depictive iconicity is considered especially unnatural in negative contexts, and can be quite natural in other, positive (especially positive episodic) contexts.

From a compositional standpoint, we might want to conclude that iconic depiction contributes differently to meaning in different semantic environments. Why might this be? One hypothesis about iconic content is that it requires existence. Kuhn (2020) argues that existence requirements on iconic content underly patterns in anaphora in sign languages. The difference between positive and negative versions of sentences in this task provide further support for this idea. In addition, we hypothesize that there may be a pragmatic component, since negation involves focus alternatives: depiction might be especially incompatible with the construction of pragmatic alternatives. Alternative semantics is type of semantic composition used to model sentence meaning in many different domains including questions, disjunction, negation, and focus. Consider, for example, that the formal semantic denotation of a question like *Did Alex eat the pasta?* is typically modeled as the set of possible answers to that question, e.g. {*Alex ate the pasta, Alex didn't eat the pasta*}. The pragmatic/semantic contribution of focus is also modeled via alternatives: the focus semantic value of *Alex ate the PASTA* (with focus on *pasta*) is typically modeled as activating a set of alternatives, e.g. {*Alex ate the pasta, Alex ate the bread, Alex ate the rice, ...*}. Focus sensitive operators like *only* or negation interact with these alternatives, so that *Alex only ate the PASTA* negates everything in the set of alternatives that isn't pasta (or isn't entailed by eating pasta), and negation with focus as in *Alex didn't eat the PASTA* seems to do the opposite, negating the alternative involving pasta and suggesting that one of the other alternatives must be true. Although merely a sketch of the way that alternative semantics is used in

compositional semantics, this should give a sense that many functional pieces of the grammar, like disjunction, questions, and negation, rely on composition with these kind of semantic/pragmatic alternatives (Rooth 2016).

Why might depiction be incompatible with alternatives? Lupyan & Winter (2018) note that symbols generalize over details in a way that more (depictively) iconic language does not. For example, a picture of a guitar necessarily shows what kind of guitar something is: it has to commit to whether it is an electric or acoustic guitar. In contrast, a symbol like the English word *guitar* or the ASL sign 'guitar' has potential to generalize over that level of detail. Such is the nature of a symbol, versus a depiction. This makes symbols ideal for creating alternatives: we can imagine the alternatives for *Alex didn't play the GUITAR* to be other kinds of instruments, e.g. {*Alex played the piano, Alex played the french horn, Alex played the guitar, ...*}. The key idea behind why depictions are incompatible with negation, questions, and other places we see alternatives is that iconic depictions are generally too specific to create convenient partitions for alternatives.

Depiction being generally incompatible with building pragmatic alternatives is consistent with quite a lot of otherwise puzzling existing data in the literature. One place we see supporting data is in the study of co-speech gestures: Ebert & Ebert (2016) focus on the way that depictive gestures seem to be unable to be targeted by negation. They attribute this to the gestural content being essentially a supplement, "not at issue", similar to a non-restrictive relative clause. Schlenker (2021) agrees with the not-at-issue status of gestures, and takes them to be a type of presupposition ("co-supposition"). Esipova (2019) also takes gesture to frequently be not-at-issue. Thus, there is a robust discusison of the difficulty of gesture to occur in the scope of negation. There are, however, two important distinctions between that work and this work that set it apart. First, that line of work generally makes the assumption that gestures are natural/acceptable in the contexts of negation, and investigates their meaning given that assumption. As we have seen, however, depictions can be rated quite unnatural in negative contexts, so we should first be verifying that they are well-formed before investigating what they mean. Second, previous work typically does not separate depictive gestures from more descriptive/lexically iconic gestures, since the gestural modality is taken to be the relevant feature, instead of the semiotic distinction (depiction vs. description). Given that the previous two sections showed lowered naturalness ratings for iconic depictions even in written language, we certainly want to keep in mind the possible variations in naturalness. Second, it seems that depiction matters, not necessarily modality, given that both of these studies were in written language, not visual/manual gesture.

Another kind of data that supports the incompatiblity of depictions and pragmatic alternatives comes from the analysis of visual pictures. Esipova (2021) cleverly analyses the compositional interaction between pictures and negative modals by investigating a small corpus of prohibition signs. These generally involve a symbol for negation (e.g. a slash through a picture) and a depicted activity. Although the intended interpretation is typically clear from context, Esipova (2021) notes that a picture with such a prohibition typically has several potential interpretations, and concludes that pictures are information structurally ambiguous. What we might

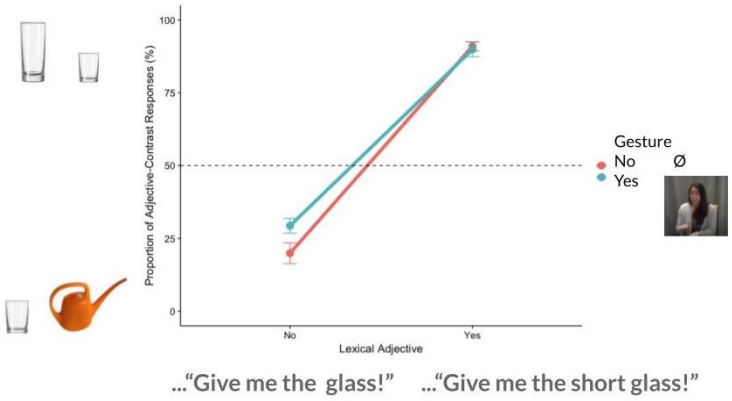


Figure 4: Proportion choosing the contrastive pair response, based on 79 participants (no exlusions) recruited from Amazon MT platform

add to this observation is that whatever negation does target, that aspect of the picture ends up having to be interpreted as a symbol that generalizes over particulars (e.g. particular kind of activities). It's unclear why this would be, except if depictions are incompatible with building pragmatic alternatives, and so any visual representation (e.g. a picture) that would naturally be interpreted as a depiction must be re-interpreted as a symbol in order to be targeted by negation, with this re-interpretation open to ambiguities.

Finally, we see evidence in prior literature for incompatibility between depiction and alternatives by studying contrastive inferences involving gesture. A well known fact at the intersection of psycholinguistics and semantics is that the use of additional modifications like asking someone to pick up *the tall glass*, versus just asking them to pick up *the glass*, leads to contrastive inferences that the modifier is necessary, i.e. there are multiple glasses, of which one is tall (Sedivy 2007). This is typically shown through real time processing measures like eye-tracking, where participants who hear *Point to the tall...* will show more eye movements toward a medium sized glass than to an even taller pitcher if there is only one pitcher but two glasses, one medium sized and one small. In other words, the use of the modifier *tall* cues the listener to guess that whatever noun follows must not be unique (*glass*) since if there was only one item (e.g. *pitcher*) then the adjective *tall* wouldn't be necessary.

Alsop *et al.* (2018) tested contrastive inferences using a simple off-line forced choice behavioral paradigm, and showed that for spoken modifiers, participants showed robust contrastive inferences. In other words, they regularly assumed that the use of an adjective like *short* indicated that the intended referent was one of a pair of items (e.g. a short and tall glass) instead of a unique item in a pair (e.g. a short glass and a watering can, see Figure 4). Intruigingly, they did not find the same effect for depictive co-speech gestures: use of a modifying gesture had a small effect in the absence of any other modifier, yet overall led to no contrastive inferences (i.e. there was no preference for the unique item, and participants were ignoring the gesture for contrastive reasoning).

Why might the depictive gestures not lead to contrastive inferences when the same information (e.g. *short*) in the spoken language did? Here again we might

attribute this to composition in alternative semantics: the gestural depiction may tell about the particular glass but it will not be taken into account in the composition of semantic alternatives.

In general, it seems that depictive co-speech gestures in English provide quite good evidence for a lack of depiction in the composition of alternatives. It is consistent with the data from the previous section that focused on written English, and from the study of visual representatins/pictures. What about other languages? The next section, Section 5, describes qualitative judgements on the naturalness of negation and depictions in spoken Japanese, American Sign Language, and English that seem to provide further cross-linguistic evidence, before Section 6 concludes.

5 Cross-linguistic and cross-modality comparison

As Dingemanse (2019) notes, depiction is perhaps easier to identify in other languages with rich arrays of depictive language. One class of languages known for rich depictive vocabulary are spoken languages with rich sets of ideophones, such as Japanese. Another class of languages known for rich depictive vocabulary are sign languages, such as American Sign Language. We will investigate each in turn here, before turning to an iconic corner of the English language.

5.1 Japanese ideophones

An especially rich discussion of the semantic contribution of Japanese ideophones, relevant to the question of negation and depiction, comes from Kita (1997). Since we will not have space to cover the details here, the reader is referred to the paper for details between different classes of ideophones, for example, that seem to vary with their syntactic category and amount of depiction. What we focus on here are a category of ideophones which are used in spoken language to support depictions. They are somewhat sound-symbolic in that the form bear a relation to their meaning: in this sense they involve lexical/descriptive iconicity. However, they can also be used depictively: Kita (1997) shows that these ideophones are far more likely to accompany depictive co-speech gestures than non-ideophones, and they are especially natural when used to both tell and show how something feels, moves, sounds, etc.

An especially relevant observation for our discussion here is that Kita (1997) shows that ideophones are resistant to use under negation. He contrasts the well formed positive sentence involving an ideophone (3) with a negative version of the same sentence (4) which is no longer acceptable. He further shows that a different (non-iconic, non-ideophone) modifier like *sizukani* 'quietly' is perfectly well formed under negation (5), illustrating that this is an incompatibility specifically between negation and ideophones.

(3) **Depiction, no negation**

tama ga	*gorogoro to korogat-ta no*	*o*	*mi-ta*
ball NOM	Mimetic roll-Past Nominalizer	ACC	see-Past

'(One) saw a ball rolled *gorogoro*.'
(*gorogoro* = movement of a heavy round object with continuous rotation)

(4) **Depiction, with negation** (not acceptable)

**tama ga gorogoro to korogat-ta no de wa na-i*
ball NOM Mimetic roll-Past Nominalizer COP Focus Neg Pres

'It was not the case that a ball rolled *gorogoro*.'

(5) **Descriptive modifier, with negation**

tama ga sizukani korogat-ta no de wa na-i
ball NOM quietly roll-Past Nominalizer COP Focus Neg Pres

'It was not the case that a ball rolled quietly.'

How might we view the Japanese pattern in light of the hypothesis that depictive material is incompatible with pragmatic alternatives? The idea is that a non-ideophone such as *sizukani* 'quietly' is an abstract symbol that generalizes over all kinds of ways that balls can roll quietly, and as such contrasts with other abstract symbols for manners of rolling, such as 'loudly', etc. In contrast, the ideophone *gorogoro* is used precisely in contexts in which the speaker wants to not only describe but also depict for their interlocutor some aspect of the ball rolling, so the speaker can't help but depict while saying the ideophone, and as such is incompatible with building alternatives.

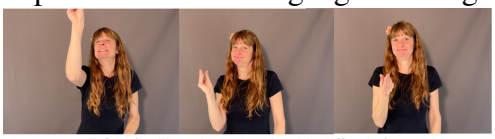
Note that this analysis assumes that a (symbolic) ideophone will generally be used in cases when one wants to (also) depict, and thus the incompatibility in the Japanese case is somewhat roundabout/indirect. In other words, there isn't strictly a prohibition on using the ideophone in a strictly symbolic/lexical sense, only when it is used to help depict; it just happens that ideophones are best used for purposes of depiction. This might actually turn out to be a strength of the analysis, as we will see when we extend the discussion to English onomatopeia in the final section. First, though, we turn to a case of more obligatory depiction: classifiers in American Sign Language.

5.2 ASL depictive classifiers

Sign languages around the world make use of a class of signs that involve conventionalized handshapes used to depict; in the formal linguistics literature the symbolic nature of these handshapes is emphasized and they are often referred to as classifier predicates (Zwitserlood 2012), since handshapes are chosen based on categories similar to the verbal classifier literature, e.g. humans, animals, vehicles, flat objects, etc. In cognitive linguistics literature, another key aspect of these signs is highlighted: the fact that their locations and movements are depictive, and thus they are known as depicting signs/depicting verbs (Dudis 2004). Like Japanese ideophones, they involve a symbolic base (the basic ideophone form in Japanese, the classifier handshape in ASL), which is most natural used to depict. In the case of ASL, the depiction is arguably even more necessary than in Japanese, as the entire purpose of this class of expressions is to depict objects, shapes, events, etc.

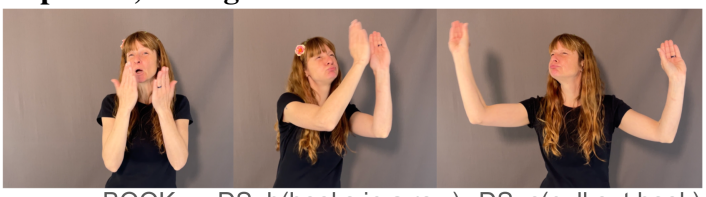
Given the fundamentally depictive nature of depictive classifier predicates (to use an especially broad terminology), and our working hypothesis in this paper that depictions are incompatible with building alternatives, we might expect to see

a similar pattern to the Japanese ideophone for ASL classifiers. And indeed, the pattern is surprisingly similar, despite the different language and language modality.


DS_c(pulling down book with difficulty)

In (6), the depicting sign is used as a predicate to depict the way that the main character pulled a book off the shelf. This sign uses the "c" handshape, conventionalized for handling objects of the size and shape of a large book. The movement shows/depicts/demonstrates the way that the book was pulled off the shelf. In this positive sentence, the depictive classifier is considered natural/acceptable by signers of ASL. However, the same sign in (7) 'pull down the book' is no longer acceptable under the scope of negation; instead, to express the idea that it was not difficult (i.e. to target the pulling-down manner with negation), an entirely symbolic sign HARD has to be used, as in (8); to maintain a depicting classifier it has to be used outside the scope of negation (as in DC_c(hold book)), and cannot show what did *not* happen, but only depict what *did* happen.

(6) **Depiction, no negation**

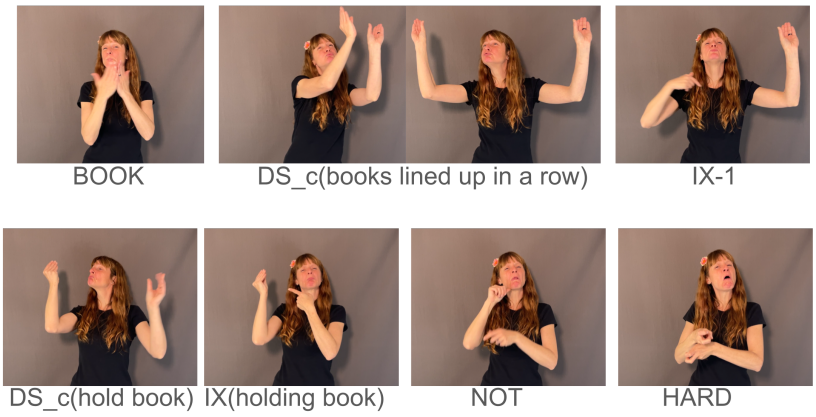

BOOK DS_b(books in a row) DS_c(pull out book)

DS_c(pulling down book with difficulty)

'Of all the books in a row, it was difficult to pull one down'

(7) **Depiction, with negation** (not acceptable)
*BOOK DS$_c$(books lined up), NOT DS_c(pull down w/difficulty)
'Of all the books in a row, it wasn't difficult to pull one down'

(8) **Descriptive modifier, with negation**



'Of all the books in a row, it wasn't difficult to pull one down'

Our very preliminary finding for the interaction of depiction and negation in ASL, then, involving classifiers, is quite similar to what we find in Japanese involving ideophones. This is not to say that this exhausts depiction in either language, of course! Much research in sign linguistics focuses on the ways that depiction interacts with other aspects of the language; see especially relevant work by Ferrara & Hodge (2018) and Hodge & Ferrara (2022), who provide a detailed discussion of different semiotic aspects of language, with an especially informed view of sign languages, that are consistent with and an inspiration for the one I am adopting here.

What I hope to emphasize as new is the way that we can test for and probe these different classes by means of the way that they interact with other parts of the grammar. For example, depictions in ASL seem to be resistant to contexts that involve negation, and also, it seems, less natural in questions. This might be precisely why we also need descriptions/symbolic language: because communication is not all about showing what happened but also about asking questions, answering in the negative, etc.

5.3 English onomatopeia

In this section so far we have shown that robust depictive language, such as Japanese ideophones or ASL depictive classifiers, are notably less acceptable in the context of negation than in positive episodic contexts, and less acceptable than symbolic language in negative contexts. It's worth noting before we leave this section that the same observations can be many in any corner of a language where we contrast depictive iconicity with symbolic/descriptive iconicity. Consider the case of English onomatopeia: in (9), a descriptively iconic symbol like the word *chirp* can quite easily support a further depiction, as in *chirrrp-chirrping[expressed in a sing-songy manner]*. This is similar to the case of the Japanese ideophone being used to depict, or the ASL classifier handshape being used in a depiction. In the case of English onomatopeia, we also see that the sentence is much less natural/acceptable under negation (10, whereas a similar example with a non-depictive modifier under negation is fine (11). Note, moreover, that we find the same contexts with a polar question (12)-(13).

(9) **Depiction, no negation**
The bird was chirrrp-chirrping[expressed in a sing-songy manner] on her perch.

(10) **Depiction, with negation** (not acceptable)
*The bird wasn't chirrrp-chirrping[expressed in a sing-songy manner] on her perch.

(11) **Descriptive modifier, with negation**
The bird wasn't chirping loudly on her perch.

(12) **Depiction, in question** (not acceptable)
*Was the bird chirrrp-chirrping[expressed in a sing-songy manner] on her perch?

(13) **Descriptive modifier, in question**
Was the bird chirping loudly on her perch?

Thus we see again that the incompatibility of negatoin and depiction isn't about iconicity in a broad sense (*chirp* is often considered iconic, from the perspective of non-arbitariness in the lexicon), or of gestures, but rather it is a property of iconic depiction in a narrow sense, including in spoken/written language. The extension of this data to questions further supports the hypothesis in Section 4 that this is about alternatives, and not, say, only about negation.

6 Discussion

We began with the question of how iconicity in language relates to compositional semantics: how does the fact that a form is iconic affect the result of its composition with another form? Does this depend on what kind of composition is involved, or what other forms it combines with? We first proposed to distinguish iconicity in the lexicon from depictive iconicity, which we argued are too frequently conflated in literature on this topic. We then narrowed our question to how depictive iconicity composes with other aspects of the language, and presented a serious of short quantitative studies that showed that depictions in written English are rated by English speaking participants as less acceptable in contexts involving negation.

The unacceptability of depiction in the context of negation was proposed to be due to the inability of depictions to build alternatives. This was further supported by existing data in co-speech gesture and picture semantics that find depictions behaving unlike symbolic language under negation and other alternative-invoking environments, such as contrastive inferences. Finally, we showed that these observations have cross-linguistic validity in rich forms of depiction in other languages such as ideophones in spoken Japanese and depictive classifiers in ASL, as well as onomatopeia in spoken English.

There are three main takeaways from this short paper. First, the semiotic approach championed by many outside of formal semantics, such as Clark (2016), Kita (1997), Dingemanse *et al.* (2015), and Ferrara & Hodge (2018) seems to provide an important insight for formal compositional semantics, since we find this resistance to negation not necessarily in iconicity originating in the lexicon, but rather in depictively iconic language. In addition, we see possible positive influence

in the other direction: alternative semantic frameworks have been extremely productive in modeling semantic composition in areas as varied as focus (Rooth 1992), questions (Hamblin 1976), disjunction (Alonso-Ovalle 2006), negation (Ramchand 1997), and broadly the organization of our entire discourse (Roberts 2012). When we look beyond positive episodic sentences that describe a particular event (as is commonly done in storytelling and narratives, and language elicited from picture prompts, etc.), we find that much of conversational language involves reasoning over pragmatic alternatives. Here we find different functions for different semiotic categories: depictions are excellent for showing the small details of particulars, but description is necessary for making generalizations of the sort involved in building alternatives for questions and their (focused) answers.

The final takeaway more narrowly concerns the formal semantic analysis of iconic content like co-speech gestures, classifier predicates, and ideophones. The gesture literature has frequently note that gestures are largely incompatible with negation. Here, we extend this observation much more broadly to depictive content, irrespective of modality. Sometimes "gesture" is used to encompass everything here that we are calling iconic depictions (Goldin-Meadow & Brentari 2017), and so under that interpretation indeed, gestures seem to be unable to be targeted by negation. But we argue that this is due to their nature as depictive, not their modality. This connects with quite a lot of other literature on the not-at-issueness of gestural/depictive content more broadly, hopefully combining a large class of literature not by their language modality, but rather their semotic status.

The ideas discussed here naturally extent to depictive content wherever it might be found. This paper covered written quotations, lengthened vowels (in writing), ideophones and onomatopeia in spoken and written language, and sign language depictive classifiers. There are as many ways to depict as there are to describe, and many of them appear together with symbolic language, so hopefully this will spur more work on the interaction of the formal semantics of functional/compositional language and iconic depictions.

Acknowledgements

Thanks to Mark Dingemanse, Masha Esipova, and the audience at the Chicago Linguistic Society 58 in Spring 2022, HSP (especially Sotoro Kita), and the joint meeting of the SPP/ESPP in Summer 2022, and members of the M&M Lab for comments on various stages of this project, and Jessica Tanner for recording ASL depictive classifier sentences. This work was supported by an NSF CAREER grant BCS-1844186.

References

Abusch, D. 2012. Applying discourse semantics and pragmatics to co-reference in picture sequences.

Abusch, D., & M. Rooth. 2022. Pictorial free perception.

Alonso-Ovalle, L. 2006. *Disjunction in alternative semantics*. University of Massachusetts Amherst.

Alsop, A., E. Stranahan, & K. Davidson. 2018. Testing contrastive inferences from suprasegmental features using offline measures. *Proceedings of the Linguistic Society of America* 3. 71–1.

Aristodemo, V., & C. Geraci. 2018. Visible degrees in Italian sign language. *Natural Language & Linguistic Theory* 36. 685–699.

Blasi, D. E., S. Wichmann, H. Hammarström, P. F. Stadler, & M. H. Christiansen. 2016. Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences* 113. 10818–10823.

Burge, T. 2018. Iconic representation: Maps, pictures, and perception. In *The map and the territory*, 79–100. Springer.

Clark, H. H. 2016. Depicting as a method of communication. *Psychological review* 123. 324.

Clark, H. H., & R. J. Gerrig. 1990. Quotations as demonstrations. *Language* 764–805.

Cohn, N. 2012. Comics, linguistics, and visual language: The past and future of a field. In *Linguistics and the Study of Comics*, 92–118. Springer.

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

Davidson, K. 2020. Is "experimental" a gradable predicate. *Proceedings of North East Linguistics Society (NELS)* 50.

Dingemanse, M. 2015. Ideophones and reduplication: Depiction, description, and the interpretation of repeated talk in discourse. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"* 39. 946–970.

Dingemanse, M. 2019. "Ideophone" as a comparative concept. *Ideophones, mimetics, expressives* 13–33.

Dingemanse, M., & K. Akita. 2017. An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese 1. *Journal of Linguistics* 53. 501–532.

Dingemanse, M., D. E. Blasi, G. Lupyan, M. H. Christiansen, & P. Monaghan. 2015. Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences* 19. 603–615.

Dudis, P. G. 2004. *Depiction of events in ASL: Conceptual integration of temporal components*. University of California, Berkeley.

Ebert, C., & C. Ebert. 2016. The semantic behavior of co-speech gestures and their role in demonstrative reference. presentation, LSCP LANGUAGE seminar talk at the Institut Jean-Nicod, Département d'études Cognitives, école normale supérieure, Paris, Dec. 2016.

Emmorey, K. 2014. Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369. 20130301.

Esipova, M. 2019. *Composition and projection in speech and gesture.*. New York, NY: New York University dissertation.

Esipova, M. 2021. On not-at-issueness in pictures. *Glossa: a journal of general linguistics* 6. 83.

Ferrara, L., & G. Hodge. 2018. Language as description, indication, and depiction. *Frontiers in Psychology* 9. 716.

Fodor, J. 2007. The revenge of the given. *Contemporary debates in philosophy of mind* 105–116.

Goldin-Meadow, S., & D. Brentari. 2017. Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences* 40.

Greenberg, G. 2021. Semantics of pictorial space. *Review of Philosophy and Psychology* 12. 847–887.

Hamblin, C. L. 1976. Questions in montague english. In *Montague grammar*, 247–259. Elsevier.

Hochgesang, J., O. Crasborn, & D. Lillo-Martin. 2020. ASL Signbank. https://ASLsignbank.haskins.yale.edu/.

Hodge, G., & L. Ferrara. 2022. Iconicity as multimodal, polysemiotic, and plurifunctional. *Frontiers in Psychology* 13.

Kita, S. 1997. Two-dimensional semantic analysis of japanese mimetics. *Linguistics* 379–415.

Kuhn, J. 2020. Logical meaning in space: Iconic biases on quantification in sign languages. *Language* 96. e320–e343.

Kuhn, J., & V. Aristodemo. 2017. Pluractionality, iconicity, and scope in French Sign Language. *Semantics and Pragmatics* 10.

Lupyan, G., & B. Winter. 2018. Language is more abstract than you think, or, why aren't languages more iconic? *Philosophical Transactions of the Royal Society B: Biological Sciences* 373. 20170137.

Marty, P., E. Chemla, & J. Sprouse. 2020. The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: a journal of general linguistics (2016-2021)* 5. 72.

Meir, I. 2010. Iconicity and metaphor: Constraints on metaphorical extension of iconic forms. *Language* 86. 865–896.

Ramchand, G. C. 1997. Questions, polarity and alternative semantics. In *North East Linguistics Society*, volume 27, p. 28.

Roberts, C. 2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and pragmatics* 5. 6–1.

Rooth, M. 1992. A theory of focus interpretation. *Natural language semantics* 1. 75–116.

Rooth, M. 2016. Alternative semantics. *The Oxford handbook of information structure* 19–40.

Saussure, F. d. 1916. *Course in general linguistics*. Columbia University Press, 2011.

Schlenker, P. 2018a. Gesture projection and cosuppositions. *Linguistics and Philosophy* 41. 295–365.

Schlenker, P. 2018b. What is super semantics? *Philosophical Perspectives* 32. 365–453.

Schlenker, P. 2021. Iconic presuppositions. *Natural Language & Linguistic Theory* 39. 215–289.

Schlenker, P., J. Lamberton, & M. Santoro. 2013. Iconic variables. *Linguistics and philosophy* 36. 91–149.

Sedivy, J. C. 2007. Implicature during real time conversation: A view from language processing research. *Philosophy compass* 2. 475–496.

Strickland, B., C. Geraci, E. Chemla, P. Schlenker, M. Kelepir, & R. Pfau. 2015. Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences* 112. 5968–5973.

Taub, S. F. 2001. *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge University Press.

Tieu, L., R. Pasternak, P. Schlenker, & E. Chemla. 2017. Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks. *Glossa: a journal of general linguistics* 2.

Tieu, L., P. Schlenker, & E. Chemla. 2019. Linguistic inferences without words. *Proceedings of the National Academy of Sciences* 116. 9796–9801.

Winter, B., & M. Perlman. 2021. Iconicity ratings really do measure iconicity, and they open a new window onto the nature of language. *Linguistics Vanguard* 7.

Zwitserlood, I. 2012. Classifiers. In *Sign language: An international handbook*. De Gruyter.