



## Super Linguistics: an introduction

Pritty Patel-Grosz<sup>1</sup> · Salvador Mascarenhas<sup>2,3,4</sup> · Emmanuel Chemla<sup>3,4,5</sup> · Philippe Schlenker<sup>2,3,4,6</sup>

Accepted: 23 November 2022 / Published online: 22 June 2023  
© The Author(s) 2023

**Abstract** We argue that formal linguistic theory, properly extended, can provide a unifying framework for diverse phenomena beyond traditional linguistic objects. We display applications to pictorial meanings, visual narratives, music, dance, animal communication, and, more abstractly, to logical and non-logical concepts in the ‘language of thought’ and reasoning. In many of these cases, a careful analysis reveals that classic linguistic notions are pervasive across these domains, such as for instance the constituency (or grouping) core principle of syntax, the use of logical variables (for object tracking), or the variety of inference types investigated in semantics/pragmatics. The aim of this overview is to show how the application of formal linguistic concepts and methodology to non-linguistic objects yields non-

---

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10988-022-09377-8>.

---

✉ Pritty Patel-Grosz  
pritty.patel-grosz@iln.uio.no  
Salvador Mascarenhas  
salvador.mascarenhas@ens.fr  
Emmanuel Chemla  
em.chemla@gmail.com  
Philippe Schlenker  
philippe.schlenker@gmail.com

- <sup>1</sup> University of Oslo, Oslo, Norway
- <sup>2</sup> Institut Jean-Nicod (ENS – EHESS – CNRS), Paris, France
- <sup>3</sup> Département d’Etudes Cognitives, Ecole Normale Supérieure, Paris, France
- <sup>4</sup> PSL Research University, Paris, France
- <sup>5</sup> LSCP (ENS – EHESS – CNRS), Paris, France
- <sup>6</sup> Department of Linguistics, New York University, New York, USA

trivial insights, thus opening the possibility of a general, precise theory of signs. (An appendix, found in the online supplements to this article, surveys applications of Super Linguistics to animal communication.)

**Keywords** Formal linguistics · Theoretical linguistics · Super Linguistics · Primate linguistics · Sign language grammar · Gestural grammar · Picture semantics · Visual narratives · Dance · Music · Iconicity · Concepts · Reasoning

## 1 Introduction

Semiotics was initially conceived as a general theory of natural signs (see Morris, 1938), be they human or animal—or neither, as in the case where spots “mean” measles (Grice, 1957: 377). The focus has largely been on their meaning, though Morris (1938: 6–7) explicitly lays out the distinction between a *syntactical*, *semantical* and *pragmatical dimension of semiosis*, in connection with the goal of achieving a general theory of signs. By contrast, formal linguistics investigates form and meaning in human language with sophisticated tools from formal language theory and logic, with rich descriptive and experimental data from numerous languages (e.g. Gutzmann et al., 2020; Maienborn et al., 2011). Contemporary formal linguistics consists of two branches: the study of form (phonology, morphology and syntax), and the study of meaning (semantics and pragmatics, broadly called ‘formal semantics’). The goals of formal linguistics, as opposed to the more general aims of semiotics, have traditionally been restricted to human language, in fact to spoken languages, followed by a gradual and relatively recent integration of sign languages phonology, then morphology and syntax, and more recently semantics (e.g. Klima and Bellugi, 1979; Pfau et al., 2012; Quer et al., 2021; Sandler and Lillo-Martin, 2006; Stokoe, 1960).

Several recent extensions of formal linguistics, which in part amount to integrating formal linguistics and semiotics into a kind of formal semiotics (see Greenberg, 2021b; Schlenker, 2018d), suggest that it might contribute to a general theory of signs as originally envisioned by Morris (1938). In so doing, formal linguistics may come to encompass aspects of human communication (such as gestures and facial expressions) that were traditionally left outside its purview, as well as non-linguistic systems such as animal communication, visual narratives, music and dance. It might even extend to reasoning and concept manipulation, where the form (the ‘language of thought’) must be inferred on indirect grounds. The aim of pursuing a general theory of signs that incorporates the tools and techniques of formal linguistics is to generate results to which formal and experimental sciences will make sustained reference in the future.

In this introduction, we sketch initial results of this broader field of ‘Super Linguistics’ (using ‘super’ in the original Latinate sense ‘beyond’, e.g. ‘beyond language’, but also ‘beyond standard objects of linguistic inquiry’), with special attention to structure and meaning phenomena. In brief, Super Linguistics makes it possible to investigate diverse representational forms with increased precision; it

offers an interesting typology of structure and meaning operations in nature; and it draws surprising new connections among these domains, including between human words, body movement and music and general reasoning abilities (an extension to animal calls is surveyed in the Appendix). Among a range of other findings, we will see that visual narratives, music and dance (the three non-linguistic objects discussed in Sect. 5) all share two properties: on the syntactic side, they exhibit mechanisms for creating grouping structure, and on the semantics side, they make use of abstract variables. Before we lay out these results, however, we must first introduce key achievements of formal linguistics; each will serve to illuminate empirical domains that belong to Super Linguistics.

## 2 Achievements of formal linguistics

The Chomskyan revolution of the 1950s and 1960s showed that human languages can be analyzed as formal languages; this was a major step towards finding rules that systematically explain how sentences of natural language are put together and convey information. The next step, pioneered by Montague (1970a, 1970b, 1973), was to treat human language as a formal *interpreted* language, and thus to systematically associate syntactic rules with semantic (i.e. meaning) rules. But what is a semantic rule? The key insight, from the philosophy of language (e.g. Davidson, 1967), was that *to know the meaning of a sentence is to know under what conditions it is true*. As a result, the key problem was to systematically associate truth conditions to sentences of arbitrary complexity. This reduced the problem to one that had been solved much earlier for (far simpler) logical languages: Tarski (1983 [1935], 1944) had shown how to define systematic truth conditions for mathematical logic. The challenge was thus to assess whether, and how, these logical methods could be extended to human languages.

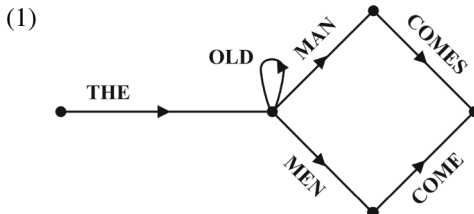
This multi-generation enterprise has yielded systematic accounts of form and meaning in human language. Four questions that have been raised in linguistics are of particular relevance for Super Linguistics: (1) where does human language fit within the hierarchy of formal languages? (syntax), (2) how is information about the world encoded in words? (semantics in the strict sense), (3) how is semantic information enriched by reasoning based on the beliefs of the speech act participants? (pragmatics), (4) how are semantic and pragmatic operations acquired and processed in real time? (psycholinguistics of meaning). We discuss these four questions, one by one, in Sects. 2.1–2.4.

### 2.1 Syntax: where does human language fit in the hierarchy of formal languages?

#### 2.1.1 From the syntax of human language to the syntax of non-standard objects

Generative syntax started with a bang with the publication of Chomsky's *Syntactic Structures* in 1957. Treating human language as a formal language, it included a result about what it could *not* be. Chomsky presented plausible predictions with

regards to which sentences are and are not acceptable in English; these showed that human language could not be generated by a finite state machine, as illustrated in (1) for a grammar with the sentences *the old man comes*, *the old men come*, *the old old old old man comes*, etc. Chomsky (1957) shows that English has properties such as center-embedding (*ab*, *aabb*, *aaabbb*, ..., or, more generally  $a^n b^n$ ), which make it impossible to derive all utterances of English by simply transitioning consecutively from one word in a sentence to the next (Chomsky, 1957: 18–23, see also Jäger and Rogers, 2012 for recent discussion).



(illustration redrawn from Chomsky, 1957: 19)

This gave rise to a long line of research that sought to establish the following, which we discuss in turn: (i) the detailed properties of natural language syntax, both in English and across languages<sup>1</sup>; and (ii) where in the hierarchy of formal languages (‘regular’ languages with finite state grammars, context-sensitive languages, etc.) natural languages lies.

Notably, when Chomsky laid out the goals of linguistic investigation in *Syntactic Structures*, human language may well have been the object of study. But the associated questions and tools (notably, the hierarchy of formal languages) have been fruitfully applied to further areas—for example, animal signals, including cases in which their formal properties are very different from human language (e.g. where they lie much lower on the ‘Chomsky hierarchy’ of formal language complexity). Similar extensions were natural in further areas: constituent analysis originated in studies of language (see Sect. 2.1.2) but turned out to be useful in the analysis of visual narratives and even music; variables originated in logic and semantics (see Sect. 2.2.2) but illuminated aspects of visual narratives.

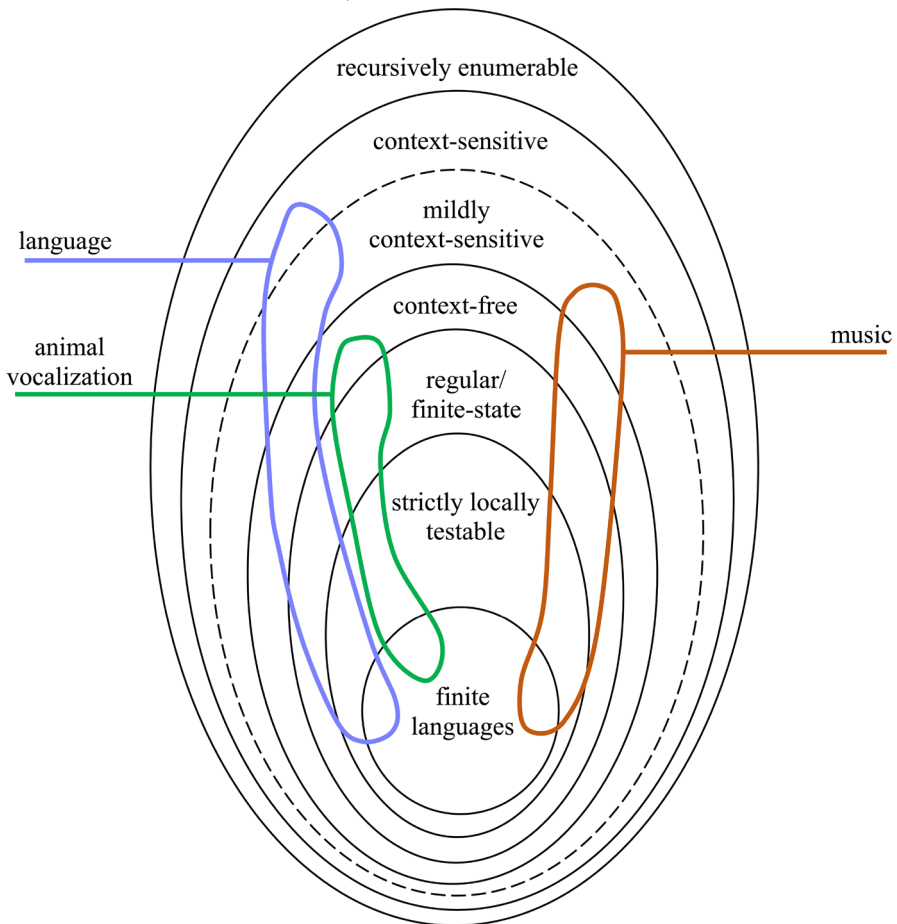
As we focus on the study of human language for a bit longer, we observe that Chomsky (1957) established an important distinction. Some insights, such as the above-mentioned argument that finite state grammars cannot generate human language, can be made solely on the basis of the set of sentences that are or are not acceptable in the language (the ‘weak generative capacity’ of a grammatical formalism). But linguists are generally interested in the question of which types of *structures* a grammatical formalism predicts (its ‘strong generative capacity’), i.e. hierarchical constituent structures, which are taken to form the basis of compositional semantic interpretation. The shift from an earlier focus on ‘weak generative capacity’ (acceptable sentences) to a later focus on ‘strong generative capacity’ (possible structures) is generally taken to be articulated in texts such as Chomsky

<sup>1</sup> See Wiltschko (2018) for a discussion on the cross-linguistic developments within this research program.

(1980: 220). This distinction is also important for the study of objects outside of natural language, as we can probe for possible/predicted structures in animal communication, music, dance, visual narratives, and so forth.<sup>2</sup>

When we compare human language to other potential objects of study, there is general agreement about the general part of the hierarchy of formal languages in which natural language syntax lies, illustrated in (2): definitely beyond finite-state (as discussed above), and certainly beyond context-free grammars (i.e. phrase-structure grammars, such as programming languages, Jäger and Rogers, 2012: 1959).

- (2) Hypothesized classification of human language, (human) music and animal vocalizations in the hierarchy of formal languages (redrawn and slightly adapted from Rohrmeier et al., 2015).



<sup>2</sup> Another object of study that will not be in the focus of this survey consists of diagrams, which have been analyzed from the perspectives of syntax and semantics, and have been discussed as early evidence for non-linguistic representations (see e.g. Shin, 1994; Shimojima, 2015, and the papers in Allwein and Barwise, 1996).

Natural languages have been argued to be within a ‘mildly context-sensitive’ region between context-free languages (where rules require one non-terminal category as their input) and context-sensitive languages (where rules can have terminals and non-terminals in their input); see for instance Jäger and Rogers (2012), Rohrmeier et al. (2015), amongst others.

As we proceed to compare natural language to other potential grammars, such as a grammar of animal communication systems, music, dance, or pictorial representation, a simplistic division between language and non-language grammars cannot be maintained. On the one hand, not all parts of language seem to require powerful grammars: phonological rules are usually thought to lie within the finite-state bound (Heinz and Idsardi, 2011). This parallels most animal communication systems that have been described, which are also well within the finite-state bound; this includes sophisticated birdsongs which we briefly discuss in the Appendix. On the other hand, there is an ongoing search for animal communication systems that clearly go beyond this bound (see e.g. Fitch and Hauser, 2004; Gentner et al., 2006; van Heijningen et al., 2009; Kershenbaum et al., 2014; Ferrigno et al., 2020).

Things are different again when one studies *human* representational systems that are non-linguistic in nature. Here, structure deriving from grouping/constituency is often one of the first lines of formal inquiry, as has been the case in the study of music, dance and visual narratives. As we will see below, sophisticated principles of structural grouping have been posited in all three objects of study; moreover, in the case of Western tonal music there are precise arguments (Rohrmeier, 2011) to the effect that music syntax goes beyond the finite-state bound, as is illustrated in (2).

However, there is one crucial difference between human language and representational systems on the one hand, and animal communication systems on the other: in the case of the latter, one has no choice but to assess grammatical formalisms on the basis of animal behavior data; this contrasts with human systems where a direct evaluation of stimuli is possible, which can be directly communicated to a researcher. For the study of human language and representational systems, researchers can generate and study the acceptability and meaning of ‘minimal pairs’, i.e. minimal modifications of one and the same object,<sup>3</sup> such as a melody in the study of music syntax, see Sect. 5.2.1.

---

<sup>3</sup> The method of ‘minimal pairs’ in formal linguistics corresponds to that of ‘controlled experiments’ in experimental fields. In both cases, what counts as ‘minimal’ or ‘controlled’ depends on an implicit theory of which properties may be relevant for the question at hand: Spanish *alguno* ‘any.masc.sg’ and *alguna* ‘any.fem.sg’ form a minimal pair for grammatical gender, but the modification also entails changes unrelated to gender, such as the phonology (change in the final vowel). The pair is ‘minimal’ only to the extent that the phonology doesn’t matter.

## 2.2 Constituency and grouping

A key notion of natural language syntax is that of constituency and grouping; for instance, in (3), the group of words *the wild deer* forms a unit with a distinct syntactic behavior, a so-called constituent. Similarly, *feed the wild deer* forms a constituent; by contrast, *feed the wild* does not.

(3) I will [feed [the wild deer]].

Constituency diagnostics allow us to determine such groups in natural language; for instance, constituents can be topicalized, as in (4)a and (4)b vs. (4)c (for the sentence in (3)), or occur in the predicative position of a pseudocleft created with the material of the sentence to be probed, as in (5)a and (5)b vs. (5)c.

- (4) a. [The wild deer], I will feed.  
 b. [Feed the wild deer], I will.  
 c. \*[Feed the wild], I will deer.
- (5) a. What *I will feed* is [the wild deer].  
 b. What *I will do* is [feed the wild deer].  
 c. \*What *I will do deer* is [feed the wild].

Syntactic grouping of this kind, amounting to constituency, can be found in many objects of study outside of natural language, including visual narratives, music and dance, as we will see below. A deeper understanding of those objects of study will thus benefit from applying the analytical tools and devices established in formal linguistic analysis.

## 2.3 Semantics: how is information about the world encoded in words?

We can now turn to the second achievement of formal linguistics in the study of human language: semantics. To investigate semantics, a key notion is that of *compositionality*, positing rules according to which the meaning of a sentence is computed on the basis of the meaning of its elementary parts (the lexical items) and the way they are put together (the syntax). Meaning is thus derived from two components, pertaining to the meaning of elementary expressions, and their mode of composition. We introduce relevant machinery by reviewing insights on constraints on possible meanings that emerge, followed by the role of variables, and compositionality itself. In the subsequent sections we will see how the study of objects outside of human language can benefit from the very same machinery.

### 2.3.1 Expressive power and constraints on possible meanings

Elementary expressions of semantics are the minimal meaning-bearing units of language; these may be closed-class (such as the morphemes ‘some’, ‘the’, ‘in’ and

‘-ed’) or open-class (such as the morphemes ‘dog’, ‘laugh’ and ‘happy’). Both cases exhibit systematic constraints on their semantic behavior. Constraints on closed-class (= grammatical/logical) expressions, such as the determiner *some* in (6)a, have been of particular interest to formal semantics. The contribution of *some* is partly similar to that of the quantifier *for some x* (written as  $\exists x$ ) in logic. The truth conditions of the English sentence in (6)a come close to those of the logical formula in (6)b, which belongs to Predicate Logic, the most common logic used to formalize mathematics and other scientific fields.

- (6) a. Some senator is honest.  
 b.  $\exists x(\text{senator}(x) \text{ and } \text{honest}(x))$

But as soon as further cases are considered, an analysis that models natural language expressions in terms of Predicate Logic falls short. *Most* as in *most senators are honest* cannot be analyzed in terms of Predicate Logic: natural language is a more expressive system, one with ‘generalized’ quantifiers—we not only find *some* and *every* as in Predicate Logic, but also *most*, *no*, *exactly one*, etc. Natural language quantifiers thus differ from quantifiers of Predicate Logic in having a more diverse repertoire. In addition, a noun phrase can also *restrict* a natural language quantifier: for instance, *most senators are honest* has an approximate meaning where there are more honest senators than there are dishonest senators (see e.g. Barwise and Cooper, 1981). This cannot be defined in terms of *most things in the universe satisfy certain properties*, which is what Predicate Logic would require if we just added the quantifier *most x* to its repertoire (e.g. Peters and Westerståhl, 2006: 473). For instance, it won’t do to treat *most senators are honest* as: *most things are senators and are honest*—an obvious falsehood; other paraphrases inspired by Predicate Logic fail as well.

As outlined above, work in formal syntax following the Chomskyan revolution sought to situate precisely the syntactic complexity of human language within a hierarchy of formal languages of increasing sophistication. The same enterprise has been conducted in logical (formal) semantics, but on the meaning side: semanticists can provide a reasonable approximation of the expressive power of human languages relative to all sorts of formal languages (e.g. Keenan and Westerståhl, 2011); as we have just seen, human language is strictly more expressive than Predicate Logic.

Moving from the expressive power of human language to the very nature of its meanings, a panoply of constraints on natural language semantics have been discussed. One is that natural language expressions differ in their entailments (captured by the terms *positive* and *negative* in Barwise and Cooper, 1981), as follows. The sentence *Sam is in Paris* entails a modified sentence generated by replacing *in Paris* with the less restrictive property *in France*: this is the hallmark of a *positive* (or ‘upward entailing’) expression (here: *Sam is* \_\_\_). By contrast, the expression *Sam is not* \_\_\_ is *negative* (or ‘downward entailing’) in that it licenses an entailment in the opposite direction: from the sentence *Sam is not in France*, one can infer that *Sam is not in Paris*, as seen in (7):



- (7) Positive operators: **Sam is** in Paris  $\Rightarrow$  **Sam is** in France  
 Negative operators: **Sam is not** in France  $\Rightarrow$  **Sam is not** in Paris

From such a perspective, a subset of natural language quantifiers can themselves be classified as positive or negative, based on how they affect a larger expression that they occur in: *some* as in *some senator* \_\_ is positive, *no* as in *no senator* \_\_ is negative, as seen in (4). This is a case where elementary grammatical (closed-class) words are positive or negative.

- (8) Positive quantifiers: **Some senator is** in Paris  $\Rightarrow$  **Some senator is** in France  
 Negative quantifiers: **No senator is** in France  $\Rightarrow$  **No senator is** in Paris

By contrast, *exactly one* \_\_ is neither positive nor negative: *exactly one archbishop is in France* does not entail that *exactly one archbishop is in Paris* (e.g. there might be zero in Paris and one in Lyons); and conversely, *Exactly one archbishop is in Paris* does not entail that *Exactly one archbishop is in France* (e.g. there might be one in Paris and another one in Lyons). Noticeably, *some* and *no* are elementary expressions but *exactly one* (two words) is not; from this restricted example one could thus see emerging a constraint according to which only a subset of quantifiers can be elementary expressions, namely the ones that can be classified as positive or negative. This indeed illustrates one of the simplest forms of a semantic universal, which puts constraints on the possible meanings of natural language elementary expressions.

While closed-class items (like quantifiers) have received particular attention, constraints on the meaning of open-class words like *dog* have been investigated as well. Gärdenfors (2004, 2014) proposes that one constraint on lexical expressions is connectedness. Formally, the denotation of a word is connected if for any two elements in it, all elements that are between those two are also in the denotation. Take the word *animal*, with the assumption that the duck and the badger are instances of animals, and thus parts of its denotation. If one has reason to think, at any level of reasoning or conceptualization, that the platypus is intermediate between ducks and badgers, one will infer by connectedness that the platypus is an animal. Interestingly, this notion has recently been unified with a constraint on quantifiers: the property of being positive (*some*) or negative (*no*) in the above sense is subsumed under the term *monotonicity* (with quantifiers of the *exactly*-type being non-monotonic). Connectedness can then be seen as a generalization of monotonicity in open-class expressions.<sup>4</sup> We will see that strikingly, these constraints have precursors in non-human animals; this shows how the application of formal linguistic tools to research on non-human animals can benefit our understanding of animal minds.

<sup>4</sup> Technically speaking, this can be formally stated as follows: a quantifier  $Q$  is monotonic if  $Q$  and its negation *not*  $Q$  are connected (Chemla et al., 2019a), and conversely a quantifier  $Q$  is connected if it is a conjunction of monotonic quantifiers (Thijssen, 1983).

### 2.3.2 Variables

When considering the merits of formal semantics, a particularly relevant aspect of the logical machinery of natural language manifests in variables: we will see that these can be put to use repeatedly in the study of objects that go beyond natural language. To begin with, variables constitute a further example in which language both resembles and differs from Predicate Logic—and both the similarities and the points of departure are enlightening. Variables, typically notated as  $x$ ,  $y$ ,  $z$  in mathematics (and Predicate Logic), are often thought to be real yet unexpressed in spoken language.<sup>5</sup> To illustrate, the natural language sentence *Sarkozy told Obama that he would be elected* is ambiguous between a reading on which *he* refers to Sarkozy and one on which *he* refers to Obama. Many linguists have argued on indirect grounds that disambiguation is effected by way of invisible variables, as illustrated in (9)a: *he* may refer to *Sarkozy<sub>x</sub>* or *Obama<sub>y</sub>*, and disambiguation is effected because *he* is mentally represented with the variable  $x$  or  $y$ . (9)b displays the same type of disambiguation, but the variables are now dependent on ('bound by') quantifiers. Variables can also be used to model cross-sentential anaphoric dependencies between an indefinite and a pronoun, (9)c.

- (9) a. **Sarkozy<sub>x</sub>** told **Obama<sub>y</sub>** that  $he_x / he_y$  would be elected.  
 b. [**A representative**]<sub>x</sub> told [**a senator**]<sub>y</sub> that  $he_x / he_y$  would be re-elected.  
 c. [**Some senator**]<sub>x</sub> is honest.  $She_x$  opposed the proposal.

Importantly, once a framework with variables was in place, semanticists uncovered some differences between natural language variables and those of Predicate Logic. In logic, a quantifier such as  $\exists x$  in  $\exists xF$  can only control variables  $x$  that appear in the immediately following formula  $F$ . This is compatible with the bound variable analysis in (9)b. By contrast, in (9)c, the quantifier [*some senator*]<sub>x</sub> controls a variable  $x$  (on *she<sub>x</sub>*) that appears in another sentence altogether, and thus cannot be modeled in Predicate Logic. To allow for such dependencies, non-standard logics called 'dynamic logics' have been developed for natural language (e.g. Heim, 1982; Kamp, 1981).

Positing abstract variables was initially motivated indirectly by their explanatory force. While the need for variables has been called into question in the semantic analysis of English (Jacobson, 1999), in line with the fact that they are not visible, variables can, remarkably, be argued to be overt in sign languages, namely as positions in signing space, called 'loci'. In (10), *past senator* co-occurs with a pointing sign (or 'index') *IX-a* towards position  $a$ , on the right from the signer's perspective, and *current senator* co-occurs with pointing sign *IX-b* towards position  $b$ , on the left. (*IX-1* is a sign where signers point towards themselves.) The two pronouns in the second sentence can then be fully disambiguated: *IX-a* refers to the former senator, *IX-b* refers to the current senator. The positions  $a$  and  $b$  play the role of the variables  $x$  and  $y$ , disambiguating the meaning of pronouns and making overt

<sup>5</sup> Note that there exist alternative, 'variable-free' analyses, for both spoken and sign language (e.g. Barker and Jacobson, 2007; Jacobson, 1999; Kuhn, 2016).

the relation between an existential quantifier and a variable that is not in its syntactic scope, and is in fact in a separate sentence, as already seen in (9)c. Not only does sign language thus arguably vindicate the use of variables: it seems to do so in the form that these have in dynamic semantics.

- (10) IX-1 KNOW [PAST SENATOR PERSON] IX-a IX-1 KNOW [NOW SENATOR PERSON] IX-b. IX-b SMART BUT IX-a NOT SMART.

'I know a former senator and I know a current senator. He [= the current senator] is smart but he [= the former senator] is not smart.' (ASL; 4, 179; Schlenker, 2014)

In addition, we will see that some loci in sign language lead a dual life, and are simultaneously logical variables and simplified pictorial denotations of what they denote. Moreover, we will see that variables have been argued to play a semantic role in visual narratives, in narrative dance, and possibly even in music; this shows how the study and understanding of such human artifacts or art forms can benefit from the application of formal linguistic tools, such as variables.

### 2.3.3 Compositionality

How can the meanings of natural language expressions be combined? This question has often received a relatively uniform answer in semantics: a combination of two expressions is interpreted by treating one expression as a function and the other as an argument.<sup>6</sup> For instance, in the sentence [*some senator*][*is honest*], [*is honest*] has as its meaning the set of people who are honest. And [*some senator*] has as its meaning a function *f* from sets *A* to truth values, with *f*(*A*) being true just in case *A* contains at least one senator.

This is a versatile framework that extends to other quantifiers, including ones that are not found in Predicate Logic. To illustrate, the meaning of *most senators* can be taken to be a function *g* for which *g*(*A*) is true just in case *A* contains more than half of the senators. What matters for our purposes is that this function-argument mechanism is very different from the mere addition of meanings by way of the pragmatic inferences we draw on the relationship between juxtaposed clauses (= concatenation), of the sort we could get for: *It's raining. The street is wet.* Here we can treat each sentence as making an independent contribution, roughly yielding a conjoined meaning such as *it's raining and the street is wet*: the two meanings can be further connected in the pragmatics without further ado (e.g. giving rise to an inference that the street is wet because it is raining). But this strategy won't normally work when function-argument combinations are involved; e.g. having established that *some senator* existentially quantifies over senators, and *is honest* denotes the set of people who are honest, it won't do to treat *some senator is honest* to mean the same as the juxtaposed clauses *There is some senator. There are honest people.* Importantly, it has been hypothesized that concatenation interpreted as conjunction suffices to combine monkey calls; see Appendix.

<sup>6</sup> For an alternative approach that aims to do away with functional application in favor of different types of conjunction-based operations, see Pietroski (2018).

## 2.4 Pragmatics: the inferential typology of language

Another achievement of formal semantics is the discovery that language doesn't just convey information, but simultaneously signals its status relative to the shared knowledge and discourse intentions of the speech-act participants. Language does so by virtue of a surprisingly rich typology of inferences, such as implicatures, presuppositions and supplements. A new question arises: What is the division of labor within this inferential typology?

### 2.4.1 *Implicatures*

To illustrate, let us go back to *Some senator is honest*. It makes two contributions: its literal meaning is that *at least one senator is honest*. But in addition, the sentence strongly suggests that *not all are*. This cannot be due to the literal meaning of *some* because it is not a contradiction to say: *Some senator is honest—in fact all are*. The inference from *some* to *not all* is defeasible, and sometimes absent (e.g. an answer *no* to *Is some senator honest?* denies that *any* is, not that *some but not all are*). A standard analysis (following Grice, 1975, see also Horn, 1972; Levinson, 2000) involves competition between *some* and *every*: the statement *every senator is honest* is more informative, hence if the speaker didn't use it, chances are that this was because she took it to be false. This inference, called an implicature, is crucially based on what we may call the Informativity Principle:

- (11) If S and S' are competing utterances, and if S' is more informative than S, prefer S' over S unless S is false.

As we will see, implicatures are found throughout human communication (including gestures). More surprisingly, implicatures (or implicature-like inferences) have been argued to explain parts of the meaning of monkey alarm calls in several species (see Appendix).

### 2.4.2 *Presuppositions*

The typology of linguistic inferences doesn't end there. *Mary stopped/continued/regretted smoking* triggers the inference that Mary smoked before. This is called a presupposition because it is presented as already established in the discourse (see for instance Beaver et al., 2021 for a recent survey), and thus strictly speaking uninformative in that it does not necessarily contribute new information to the utterance from which it arises; exceptions to uninformativity are cases where presuppositions introduce new information that can be readily accommodated by the addressee. Unlike the 'at-issue' entailment from *Mary is in Paris* to *Mary is in*

France,<sup>7</sup> the presupposition is preserved in questions and in negative statements, as shown in (12):

- (12) a. Did Mary stop/continue/regret smoking?      => Mary smoked before  
 b. Mary did not stop/continue/regret smoking.      => Mary smoked before

Presuppositions are ubiquitous. If I ask about someone in the distance: *Is she approaching?*, I am asking whether the person is approaching but presupposing that this person is female (unlike in the question: *Is this a woman approaching?*). Thus the gender specifications of pronouns generate presuppositions (see Cooper, 1983). While presuppositions are often thought to be hard-wired in the meaning of words, we will see that new gestures and even visual animations can generate presuppositions. This suggests that a productive rule lies at their source: presuppositions need not be encoded in words. Notably, while we have already highlighted cases where non-standard objects such as animal communication, music or dance benefit from the application of formal linguistic tools, this is a clear case where formal linguistic theory benefits directly from the inclusion of non-standard objects.

### 2.4.3 Supplements

A fourth inferential type pertains to supplements, which are the inferences typically triggered by non-restrictive relative clauses (see e.g. Potts, 2005; Schlenker, 2021a), (13a). Unlike presuppositions, they are generally presented as informative. But unlike ‘at-issue’ entailments, they trigger the same inferences as independent clauses even when they are embedded under logical operators. They thus contrast with embedded conjunctions, (13c), hence (13)a,b but not (13)c trigger the inference that *if Ann lifts weights, this will adversely affect her health*.

- (13) a. If Ann lifts weights, which will adversely affect her health, we should talk to her.  
 b. If Ann lifts weights, we should talk to her. This will adversely affect her health.  
 c. If Ann lifts weights and this adversely affects her health, we should talk to her.

We will see that supplements too can be generated with iconic gestures and visual animations, provided that they follow the words they specify.

## 2.5 Psycholinguistics of meaning: processing and acquiring meaning

The last question concerns the mental reality of semantic operations, i.e. how they are actually processed in language production and comprehension. This can in turn

<sup>7</sup> We use the term ‘at-issue’ as a synonym for ‘assertive’, in particular by contrast with ‘presuppositional’, ‘supplemental’, etc. (see Simons, 2007; Potts, 2015: 168 for related uses). We do not adopt a technical (QUD-based) definition of ‘at-issueness’, as proposed, for example, by Beaver et al. (2017: 280).

inform our study of language acquisition by children. In its early days, formal semantics repudiated questions of this cognitive sort (Thomason, 1974); however, later developments disavowed such a position. The case of implicatures offers an enlightening case study. Their formal description, going back to Grice (1975), suggests that they should be an add-on to the literal meaning of words, arising from considerations of informativity. This could be expected to incur a slight processing delay, and experimental studies have shown that this is indeed the case (e.g. Bott and Noveck, 2004). In addition, their derivation is expected to require a comparison between an utterance and its competitors, and the search for competitors might be expected to result in further processing effort. This arguably accounts for an important finding as we turn to language acquisition: when given an explicit choice among alternatives, young children avoid the under-informative one (Chierchia et al., 2001); nevertheless, they often fail to compute implicatures where adults do, which can be attributed to the additional processing cost of doing so. The key lesson is that semantic theories make predictions about language processing and acquisition, highlighting the fact that semantics is part of cognitive science. We return to this conclusion in Sect. 6, where we find that a radical extension of this cognitive interpretation of semantics can give rise to a study of concepts and of reasoning.

## 2.6 Interim summary

In sum, contemporary linguistics offers a precise analysis of the form and meaning of natural language expressions, treating language both as a formal and as a cognitive system. One key question on the syntactic side (Sect. 2.1) is where natural language lies in the hierarchy of formal languages; going beyond natural language, the same question can and should be asked about systems such as animal call sequences and human artifacts such as dance, music, and visual narratives. One key question on the meaning side is which aspects of semantic information are encoded in words (Sect. 2.2), and which are due to productive rules (Sect. 2.3)—a candidate for a productive rule being, as we will see, the algorithm for presupposition generation (see Sect. 4.2.2). A strategy for addressing the question of whether meanings arise from productive rules is to investigate new, sometimes invented meaning-bearing forms (such as novel gestures and visual animations); with such objects, conventionalization can be excluded as an explanation for how the inferences arise. More broadly, we would also want to determine how the different semantic operations compare to those present in other semantic systems in nature, and what we can learn from the comparison.

## 3 Integrating logical and iconic semantics

Iconicity is an important way of producing meaning in spoken and especially signed languages. Informally, *iconicity* can be defined as resemblance-based meaning; in other words, an iconic sign resembles or imitates that which it denotes. Greenberg (2021b) generalizes this idea, proposing that a semantics is iconic whenever the *form* of the individual sign plays an essential role in computing its

meaning/denotation; in an adaptation of Greenberg's own example, a fuel level gauge has an iconic semantics because the *angle* of the needle on the dial (which is part of its form) is mapped to the *amount* of fuel in the tank (which is part of its meaning).

(14) **Iconic semantics:** fuel level gauge



Picture with CC Licence

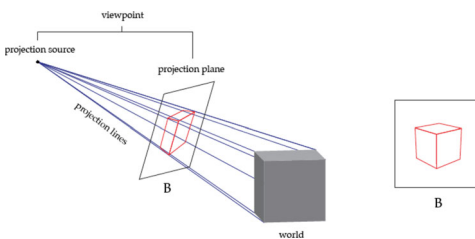
<https://commons.wikimedia.org/wiki/File:Tankanzeige.jpg>

A key challenge is to understand, both empirically and formally, how iconic and symbolic (convention-based) semantics interact; this will have benefits for the analysis of meaning in speech, signs and gestures.

### 3.1 Pictorial semantics

Pioneering work by Greenberg (2013, 2021a, 2021b) on the formal semantics of pictures (including drawings, paintings, etc.) elucidates the workings of pictorial iconicity. The semantic content of a picture is obtained by asking which situations could give rise to the markings found on the picture based on the projection method used. One projection method among several is perspective projection, illustrated in (15).<sup>8</sup>

(15) Perspective projection (Greenberg, 2021a)



<sup>8</sup> Note that there are different systems of iconic semantics; acoustic iconicity (e.g. in onomatopoeias or in music semantics) would have to be handled by different principles from perspective projection in picture semantics.

This naturally gives rise to a definition of truth for pictures: a picture  $P$  is true of those situations that can project onto  $P$  relative to the viewpoint and the system of projection used.

- (16) A picture  $P$  is true in situation  $w$  relative to viewpoint  $v$  along the system of projection  $S$  iff  $w$  projects to  $P$  from viewpoint  $v$  along  $S$ , abbreviated as:  $\text{proj}_S(w, v) = P$ .

Abusch (2013, 2020) uses this framework to account for the semantics of silent comics, a point to which we return in Sect. 5.1.<sup>9</sup> But iconic semantics also interacts in interesting ways with symbolic semantics in sign languages, to which we now turn.

### 3.2 Logical visibility and iconicity in sign language

Just like spoken languages, sign languages are traditional objects of linguistic study. Sign languages have been shown to display the same general grammatical and semantic properties as spoken languages (e.g. Sandler and Lillo-Martin, 2006), while sometimes expressing aspects of Logical Forms in a more transparent fashion, as we saw in Sect. 2.2.2. But they also make greater use of iconicity than spoken languages, for instance to modulate the form of conventional words, and iconicity per se is a non-standard object within formal linguistics, which thus falls under ‘Super Linguistics’.<sup>10</sup> A simple example is the verb *GROW* in ASL, which is (conventionally) realized as shown by the pictures in (17). But it can also be modulated in iconic ways by making the endpoints more or less broad, and by realizing the sign more or less quickly, with unmistakable semantic effects: the broader the end points of the sign, the larger the final size of the group; and the more rapid the movement, the quicker the growth process. These effects instantiate gradience, a recurrent property of iconic meanings: a small change in the form of a sign (here: its size or the speed at which is performed) entails a corresponding change in the meaning (see e.g. Dingemanse, 2015: 950).

<sup>9</sup> Abusch (2013) uses methods from Discourse Representation Theory for implementing her proposal (see e.g. Kamp and Reyle, 1993). Relatedly, it is worth mentioning the research tradition that focuses on discourse coherence by virtue of rhetorical relations (see e.g. Hobbs, 1985, 1990; Asher and Lascarides, 2003, and Kehler, 2002). Based on a multi-modal construal of discourse coherence and rhetorical relations, this research tradition has been fruitfully applied to super linguistic objects of study, including: gestures (Lascarides and Stone 2009a, 2009b), captions, arrows, and labeling (Alikhani and Stone, 2018a, 2018b, 2019), and pictorial animations, namely film (Bateman and Schmidt, 2013; Cumming et al., 2017; Wildfeuer, 2014).

<sup>10</sup> It should be noted other traditions do give iconicity its dues (e.g. Liddell, 2003 for sign language). But the integration of iconicity within formal linguistics (especially formal semantics) is definitely a very recent endeavor.



## (17) POSS-1 GROUP GROW\_



‘My group has been growing.’ (ASL, 8, 263; Schlenker et al., 2013)

This phenomenon is not unique to sign language, however: the adjective *long* can be modulated by lengthening the vowel to evoke a greater duration (and doing the same thing with *short* is... odd). Here the rule of iconic modulation seems to be that *the longer the vowel, the greater the duration*. Due to the representational possibilities of the signed modality, these iconic modulations are far richer in sign than in speech. In addition, it has been suggested that for *GROW* and *long* alike, iconic modulations make an at-issue contribution. Thus *loooong* means very much the same thing as *very long*, and this is the content that gets denied by a negation, as in: *The talk wasn't loooong* (see Schlenker, 2018b for ASL *GROW*).

Iconicity doesn't just interact with lexical material in sign languages: grammatical expressions such as pronouns can be modulated in a highly iconic fashion as well. Just like the feminine specification of *she* triggers a presupposition that the denoted person is female, pronouns in ASL (American Sign Language) and LSF (French Sign Language) can have *high* specifications (realized by pointing upwards) that trigger the presupposition that the denoted person is tall, powerful or important. Furthermore, these modulations can display ‘iconicity in action’: one typically points towards the part of the representation corresponding to the head, with the result that when the denoted individual is presented as rotated in various position, the loci get rotated as well. The conclusion is that sign language loci can simultaneously be variables and simplified pictures of their denotations (see Liddell, 2003; Schlenker et al., 2013; Schlenker, 2014, 2018c).

A striking case of interaction between grammar, logical semantics and iconicity involves repetition-based plurals. In ASL, one can optionally realize a plural, such as that of the word *TROPHY*, by repeating the noun. This yields standard readings of English-style plurals. But simultaneously, the shape of the repetitions can provide iconic information about the arrangement of the denoted objects (Schlenker and Lamberton, 2019): one may repeat *TROPHY* in a horizontal or triangular shape to signify that the trophies are arranged on a line or as triangle. But there is more: the repetitions can be easy to count, with clear separations between them (‘punctuated’), or hard to count, without clear separations (‘unpunctuated’). In the punctuated case (*TROPHY TROPHY TROPHY*), they typically denote as many objects as there are iterations. In the unpunctuated case, three iterations that lack clear separations (and are thus hard to count), which we transcribe as *TROPHY-rep3*, may mean *several* (typically at least three) *trophies*. Strikingly, the existence of repetition-based plurals as well as the distinction between punctuated and unpunctuated ones exists in several sign languages; moreover, it was also described in a homesigner (deaf person who grew up without access to sign language) who appeared to have invented this device, as it was absent from the gestures of his hearing mother

(Coppola et al., 2013). In addition, hearing non-signers might also understand related distinctions in repeated gestures (Schlenker, 2020), indicating a more general cognitive strategy. Schlenker and Lamberton (2022) propose a modular analysis in which the iconic nature of repetition-based plurals as well as the distinction between punctuated and unpunctuated repetitions results from the interaction between four modules: (i) logical grammar makes available a mechanism of existential quantification over pluralities; (ii) pictorial semantics allows the arrangement and number of the repetitions to iconically represent various pluralities of objects; (iii) pictorial semantics can be made blurry so as to create instances of pictorial vagueness, with the result that in unpunctuated repetitions the number of depicted objects is vague; and finally (iv) pictorial vagueness can be exploited pragmatically so as to yield ‘at least’ readings (akin to standard plurals) for these unpunctuated repetitions.

In sum, the interplay between logical semantics and iconicity is a key issue in sign language semantics/pragmatics (aspects of spoken language also benefit from iconic analyses, e.g. Dingemanse, 2013). Strikingly, since sign languages have the same general grammatical and logical resources as spoken languages, but make greater use of iconicity, sign is along some dimensions more expressive than speech.<sup>11</sup>

### 3.3 Logical and grammatical structures in gestures

Gestures offer a prime example of iconicity; we can easily reproduce the fuel level gauge example with a gesture for how high the water was during a flood (where the angle of a stretched out arm corresponds to the height of the water). In addition, however, while they have nothing like the sophisticated grammar of sign languages, gestures sometimes have a proto-grammar that is reminiscent of sign language, e.g. a gesture that describes a telic event with an endpoint may differ from a gesture that describes an atelic event without an endpoint.<sup>12</sup> In performing gestures in such a way, non-signers sometimes appear to know constraints that track some sign language rules standardly classified as ‘grammatical’, and play a particularly important role in linguistic semantics.<sup>13</sup> In order to abstract away from the special semantic issues raised by co-speech gestures, which may be in some ways parasitic on the spoken expressions they enrich (a point we revisit in Sect. 4.1), we focus for the moment on gestures that fully replace some words (henceforth ‘pro-speech gestures’).

<sup>11</sup> We use the term *expressive* in the sense of ‘expressive power’ (see Sect. 2.2.1), not in the more narrow sense of Potts (2005), where *expressive* meanings are contrasted with *descriptive* meanings. Note that a precise argument about the comparative expressive power of languages with and without iconicity has not, to our knowledge, been developed, hence our cautionary use of ‘along some dimensions’ in the text.

<sup>12</sup> Our discussion follows Schlenker (2020).

<sup>13</sup> Needless to say, this implies in no way that sign languages are ‘merely’ gestural: their sophisticated grammars have been described in great detail, and share multiple properties with those of spoken languages (see Sandler and Lillo-Martin, 2006). Rather, the argument is that *despite* their expressive limitations, gestures have a proto-grammar reminiscent of sign language.

Two examples will make this line of research concrete. Consider (18), an example of spoken language (English) with gestures in a non-signer. The expression *a mathematician* is pronounced simultaneously with an open hand (palm up) on the right (transcribed as  $\llbracket X\text{-hand-a}$ , and preceding the co-occurring expression in the transcription, which is boldfaced); similarly, the expression *a sociologist* co-occurs with an open hand on the left (transcribed as  $\llbracket X\text{-hand-b}$ ). With these loci in place, a pointing gesture can fully replace a pronoun that would be expected after *pick*: a pointing gesture or ‘index’ towards the right (transcribed as  $\llbracket X\text{-a}$ ) serves to refer to the mathematician, associated with locus *a*.

- (18) Whenever I can hire  $\llbracket X\text{-hand-a}$  [**a mathematician**] or  $\llbracket X\text{-hand-b}$  [**a sociologist**], I pick  $\llbracket X\text{-a}$ .

*Meaning:* whenever I can hire a mathematician or a sociologist, I pick the former.

Importantly, if we were to remove the gestures from (18) and translate the pointing gesture  $\llbracket X\text{-a}$  into the spoken English words *him* or *her*, *him/her* could be ambiguous between the two antecedents; by contrast, the pointing gesture is not: the gesture is not just a code for a pronoun, it also disambiguates the referent.

We can turn to other cases where sign language grammar can inform our study of gesture semantics. A second case involves certain ASL verbs where a movement from one locus towards another locus is part of their realization. For instance, *I give you* is realized with a movement going from the signer to the addressee, which is transcribed as *I-GIVE-2*. By contrast, *I give him* starts from the signer’s position and targets a third person locus, for instance *a* on the right—in which case it is transcribed as *I-GIVE-a*. Such verbs are often called ‘agreement verbs’, as the movement from one locus to another appears to instantiate subject and object agreement. The respective loci have thus been argued to display the behavior of agreement markers, although alternative analyses have been offered as well (Lidell, 2003; Lillo-Martin and Meier, 2011; Pfau et al., 2018, and Schembri et al., 2018). Irrespective of the final analysis, these agreement verbs appear to have gestural counterparts (Schlenker and Chemla, 2018), possibly reflecting a general strategy for tracking and disambiguating referents in the visual modality (Patel-Grosz et al., 2022).

To have a point of comparison, let us consider the ASL examples in (19), constructed around the agreement verb *I-GIVE-2* or *I-GIVE-a*. As we alternate in discussing gestures and sign language, a reader is advised to bear in mind that words in all-caps are glosses for individual signs of sign language, (19), or (using a different font) for speech-accompanying gestures, in (18) and (21), whereas lower case, e.g. in (18) and (21), marks words of spoken language. In the first sentence of (19), the verb *GIVE* displays object agreement with a third person locus *a*, corresponding to the younger brother, hence: *I-GIVE-a*. (This example does not require a formal introduction of the *a* locus.) Now the continuation in (19)a involves a missing Verb Phrase, which in spoken and sign language alike usually entails copying an antecedent. But something happens in the copying process: the

agreement markers can be disregarded, which explains why the continuation (19)a is acceptable even though that in (19)b (with overt copying) isn't.

(19) POSS-2 YOUNG BROTHER MONEY IX-1 **1-GIVE-a**.

'Your younger brother, I would give money to.

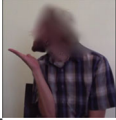
- a. IX-2 IX-1 NOT.  
You, I wouldn't.'
- b. \*IX-2 IX-1 NOT **1-GIVE-a**.


In this respect, loci are once again somewhat reminiscent of gender markers (and certain other grammatical features). Similar rules of partial copying under ellipsis are attested in English, as in (20), where the third person features and the feminine features of *her* are disregarded under ellipsis, and thus *I did too* has a reading that *I too did her homework* lacks.

(20) [Uttered by a male speaker] In my study group, Mary did her homework, and I did too.

*can mean*: I too did **my** homework.

The ASL contrasts can be replicated with gestural verbs that occur with spoken English, as seen in (21). As in ASL, a movement towards the side has to correspond to a third person object, and thus the second clause of (21)b is deviant because the object is second person but the object agreement is third person (the kisses ought to be sent towards the addressee). Strikingly, this problem doesn't arise in (21)a: as the missing verb is copied, its third person object agreement is disregarded, just as is the case for object agreement in ASL. Importantly, there are no comparable cases of object agreement in English: subjects who provided judgments on these sentences had to infer an ASL-style rule on the basis of no or extremely limited evidence.

(21) a. Your brother, I am gonna SEND-KISSES-3\_  , then you, too.

b. \*Your brother, I am gonna PUNCH-3\_  , then you, I am gonna

SEND-KISSES-3\_ 

It is noteworthy that homesigners, who grow up without access to sign language, develop gestural languages that share some properties of sign languages, but are expressively far more limited (e.g. Abner et al., 2015; Goldin-Meadow, 2003). In some cases, the reason homesigners discover such properties on their own might be

that, more generally, non-signers ‘know’ them; a case can for instance be made for repetition-based plurals, discussed above, which (i) arise without input in some homesigners, and (ii) might be intuitively known by non-signers perceiving gestures (see Coppola et al., 2013; Schlenker and Lamberton, 2022).<sup>14</sup>

An important question for future research is to determine how such instances of spontaneous grammatical learning of sign language features by non-signers can arise. Proponents of a Universal Grammar (UG) approach to human language (see e.g. Tsoulas, 2017 for a recent survey) may go as far as to argue that UG does not just specify grammatical rules in an abstract form, but also specifies part of the mapping between forms and grammatical/semantic content: a pointing sign/gesture might thus be intrinsically endowed with pronominal properties. A possible alternative view is that some signs/gestures are naturally associated with a fixed grammatical/semantic component for deeper cognitive reasons. This debate is currently open.

#### 4 Linguistics beyond words I: iconic elements within the inferential typology

Having established iconic semantics as an important area of super linguistic inquiry, we turn to the place of iconic meanings in the typology of linguistic inferences (i.e. implicatures, presuppositions, supplements, etc.). Here an important distinction must be drawn between iconic enrichments, which modify the meaning of an expression, and iconic replacements, which fully replace a word that is necessary to make the message complete.<sup>15</sup> As we will see, depending on their nature (e.g. iconic lengthening of a vowel or addition of a co-speech gesture), iconic enrichments make at-issue or non-at-issue contributions, along familiar lines of the inferential typology. Iconic replacements make a different but powerful point: their specific informational content is productively *divided* among slots of this typology (for instance with the emergence of an at-issue and of a presuppositional component); we shall see that this provides new insights about the cognitive origin of this typology by showing that inferences arise productively from a range of different objects.

To introduce the typological issue, let us consider the examples in (22), in connection with which we also introduce relevant terminology. A methodologically fruitful approach to exploring the inferential typology is to combine natural language expressions (e.g. speech) and objects with iconic meaning (e.g. gesture). Non-exhaustively, the iconic object, here instantiated by a slapping gesture, can co-occur with the natural language expression, (22)a, follow it, (22)b, or replace it, (22)

<sup>14</sup> It has also been argued that several further properties of sign language grammar can be replicated with pro-speech gestures (Schlenker, 2020). If confirmed, these results might be important to understand the historical origins of sign languages.


<sup>15</sup> We use the term ‘replace’ in a purely distributional fashion, without claiming that the corresponding mental representations involve a word that is then eliminated.

c. For the case of gestures, we use the terms *co-speech*, *post-speech* and *pro-speech*, respectively. The co-speech gesture in (22)a co-occurs with the verb *punish* (in bold type). The post-speech gesture in (22)b appears after the Verb Phrase it modifies (marked by the dash).<sup>16</sup> In (22)c, the pro-speech gesture fully replaces the verb. Finally, example (22)d illustrates something that cannot be shown by way of gesture, namely a case where the object with iconic meaning (here: vowel lengthening) modifies the natural language expression rather than merely accompanying it. Here, a conventional word, *long*, is modified in an iconic fashion by way of an ‘iconic modulation’ (which by definition is always the modification of a conventional form).

- (22) a. Co-speech gestures (co-occur with the word they modify [boldfaced])

Asterix will  **punish** his enemy.

- b. Post-speech gestures (follow the word they modify)

Asterix will punish his enemy — .

- c. Pro-speech gestures (replace a word)

His enemy, Asterix will .

- d. Iconic modulations (modify the form of a conventional word)  
The talk was loooong.

The terminology introduced in (22) is extended to sign language by replacing *-speech* with *-sign* in (22)a-c (especially *co-sign* and *post-sign* gestures/facial expressions). In Sect. 4.1, we start by considering co-speech gestures, iconic modulations and post-speech gestures, which are iconic enrichments.<sup>17</sup> In Sect. 4.2, we turn to pro-speech gestures, which are iconic replacements.

#### 4.1 Iconic enrichments in the typology of linguistic inferences: sign with iconicity vs. speech with gestures

The study of iconic enrichments (co-speech and post-speech gestures, and iconic modulations) was partly motivated by a broader assessment of our earlier conclusion that sign languages are, along some dimensions, more expressive than

<sup>16</sup> See Esipova (2019) for a broader discussion.

<sup>17</sup> As we will see, post-speech gestures behave semantically as non-restrictive relative clauses, and could thus be taken to be gestures that replace certain words or phrases that are optional.

spoken languages due to their greater use of iconicity. The question was raised if this conclusion could be premature by failing to take into account the means of iconic enrichment afforded to speech by gestures: in such a view, sign with iconicity should be compared to speech with gestures rather than to speech alone (Goldin-Meadow and Brentari, 2017). But *even* when gestures are taken into account, systematic differences arguably remain between sign with iconicity and speech with gestures. This connects to the previously mentioned inferential typology. The most salient means of iconic enrichment in speech lies in co-speech gestures. But in formal semantics approaches, it was argued from the start (notably in pioneering work by Ebert and Ebert, 2014) that co-speech gestures do not typically make at-issue contributions; this sets them apart from several iconic modulations of signs that do seem to make at-issue contributions. We already mentioned this point in connection with *GROW* above; and the shape modulations of repetition-based plurals apparently make at-issue contributions as well (Schlenker and Lamberton, 2019).<sup>18</sup>

Having established that gestures have an iconic component reminiscent of iconic signs, the key question is *how* these iconic enrichments are distributed across the inferential typology. As we will now see, none of the gestures in (22)a-c have quite the same properties as iconic modulations of the type that we find in sign language. *Co-speech gestures* have been shown to behave like triggers of presuppositions of a particular sort (called ‘cosuppositions’, sometimes used as a mnemonic for ‘conditionalized presuppositions’<sup>19</sup>), and thus fail to be at-issue, unlike many iconic modulations found in sign language. By contrast, *post-speech gestures* trigger ‘supplements’, which is the type of meaning encoded by non-restrictive relative clauses (Potts, 2005); for this reason, they too fail to be at-issue. *Pro-speech gestures*, for their part, make at-issue contributions, but unlike signs (including ones with iconic modulations) they are not conventional words at all, and are thus expressively limited for other reasons. The typology is illustrated in (23), and it turns out to be crucial to answer (plausibly in the negative, as argued in Schlenker, 2018b) the question whether speech with gestures ‘equals’ sign with iconicity.<sup>20</sup>








<sup>18</sup> See Kuhn and Aristodemo (2017) for an argument that the iconic component of repetition-based plurals can also be at-issue.

<sup>19</sup> Note that the use of ‘cosupposition’ as mnemonic for ‘conditionalized presupposition’ was introduced by Tieu et al. (2017: 6). In Schlenker (2018a: 297) (written before Tieu et al., 2017), ‘co’ in ‘cosuppositions’ was justified differently, reflecting the technical implementation of the analysis: the terminology was “intended to suggest that a cosupposition triggered in a local context *c*’ is computed in tandem with (‘co’) an at-issue component in *c*’ (by contrast, a standard presupposition triggered in *c*’ spsdown is computed before (‘pre’) any at-issue component in *c*)”.




<sup>20</sup> We do not seek to discuss here the sources of this typology; see for instance Ebert and Ebert (2014), Schlenker (2018b), and especially Esipova (2019).

Note that the disgusted facial expressions in the first and second cells of the last row of (23) are co-sign and post-sign gestures, which are not signs themselves.

(23) **Typology of iconic enrichments and replacements** (after Schlenker, 2018b)

	<b>Co-speech/co-sign gestures</b>	<b>Post-speech/post-sign gestures</b>	<b>Iconic modulations</b>	<b>Pro-speech/pro-sign gestures</b>
<b>Meaning</b>	<b>cosuppositions</b> (= presuppositions of a special sort)	<b>supplements</b> (like non-restrictive relative clauses)	<b>at-issue or not</b> , depending on the case	<b>at-issue</b> , with an additional non-at- issue component in some cases
<b>Speech</b>	Asterix will  punish his enemy.	Asterix will  punish his enemy –	The talk was loooong.	His enemy, Asterix  is going to
<b>Sign</b>	IX-arc-b NEVER  [SPEND MONEY].	IX-arc-b NEVER  SPEND MONEY]b –	POSS-1 GROUP   GROW_	[currently unclear]

To illustrate this typology, we note that the iconic enrichments in the positive sentences in (21) display radically different behaviors under negation, as seen in (24). The first thing a reader may notice is the ill-formed status of the post-speech gesture in (24)b (contrasting with (21)b). We will now go over the differences among gesture types in detail.

- (24) a. Asterix won't  punish his enemy.  
=> if Asterix were to punish his enemy, slapping would be involved
- b. #Asterix won't punish his enemy – 
- c. His enemy, Asterix won't 
- d. The talk wasn't loooong.

Most centrally, co-speech gestures and post-speech gestures differ in their acceptability when they combine with a Verb Phrase in the scope of negation.

(i) The co-speech gesture in (24)a is acceptable when accompanying a Verb Phrase that is in the scope of negation; assuming that the gesture itself is in the scope of negation, it triggers an inference that gets inherited across negation, to the effect that *if Asterix were to punish his enemy, slapping would be involved*. Crucially, further tests suggest that this inference behaves like a presupposition (for instance,



embedding under *none of the girls* will give rise to a universal inference that holds for all of the girls); as mentioned, it has received a special name (cosupposition) because the inference is conditionalized on the meaning of the modified expression (here: *punish*). Experimental data (Tieu et al., 2017, 2018) have buttressed this conclusion by means of truth value judgment tasks and (more clearly) inferential judgment tasks.

(ii) By contrast, the post-speech gesture in (24)b is simply deviant after a negated Verb Phrase. Recent literature (Schlenker, 2018b) has argued that this is because the post-speech gesture behaves like a non-restrictive relative clause and thus contributes a supplement, which is often deviant in such negative environments, as in (25).<sup>21</sup>

(25) #Asterix won't punish his enemy, which will involve slapping him.

Consideration of further examples strengthens the similarity with non-restrictive relative clauses: (26)a behaves like (26)b in suggesting that Asterix's action will involve slapping, unlike the control with a conjunction in (26)c.<sup>22</sup> Such inferential contrasts were established with experimental means in Tieu et al., 2019, and they extend to cases in which a visual animation replaces the post-speech gesture.

<sup>21</sup> A reader may wonder about the intuition that (25) can be improved by changing the mood in the non-restrictive relative clause to the counterfactual, (i).

(i) Asterix won't punish his enemy, which would involve slapping him.

Counterfactual mood has the general property of changing anaphoric properties, which is not limited to non-restrictive relative clauses, but carries over to juxtaposed main clauses, as seen in (ii) vs. (iii), or in the more standard examples from the literature in (iv), involving indefinite antecedents.

(ii) #Asterix won't punish his enemy. This will involve slapping him.

(iii) Asterix won't punish his enemy. This would involve slapping him.

(iv) Fred didn't buy a microwave oven.

a. He wouldn't know what to do with it.

b. #He might like it.


(Frank 1996: ix)

The parallels between (24)b and (25) are thus unaffected by the contrast between (25) and (i), as long as we make the plausible assumption that post-speech gestures do not come with counterfactual mood; this is plausibly due to a more general reason, namely that counterfactual mood cannot remain unexpressed.

<sup>22</sup> These inferential facts matter because the deviance of (24)b could be analogized to (i) below rather than to (25). But in view of the difference between (26)a and (26)b, the analogy with (i) seems wrong-headed.

(i) #Asterix won't punish his enemy, and this will involve slapping him.



- (26) a. If Asterix punishes his enemy — , I might scream.  
=> if Asterix punishes his enemy, slapping will be involved
- b. If Asterix punishes his enemy, which will involve slapping him, I might scream.  
=> if Asterix punishes his enemy, slapping will be involved
- c. If Asterix punishes his enemy and this involves slapping him, I might scream.  
≠> if Asterix punishes his enemy, slapping will be involved

(iii) The pro-speech gesture in (24)c makes an at-issue contribution and yields neither a cosupposition nor a supplement. Importantly, however, pro-speech gestures are expressively limited because they are not based on conventional words, unlike the iconic modulations found in sign language: in LSF, the meaning of *UNDERSTAND* or *REFLECT* can be modulated by altering the speed with which part of the sign is realized (e.g. to indicate a quick or difficult understanding or reflection, Schlenker, 2018c). It seems hopeless to represent modulations of such abstract concepts with pure gestures.

(iv) The iconic modulation in (24)d makes an at-issue contribution and triggers no conditionalized inference akin to cosuppositions.

Importantly, with the exception of pro-sign gestures (i.e. sign-replacing gestures, whose existence and status is still somewhat unclear), the same typology might hold in sign (Schlenker, 2018b). A disgusted facial expression co-occurring with the Verb Phrase *SPEND MONEY* (as on the last line of (23)) yields the same cosuppositional behavior as the slapping gesture co-occurring with *punish*: ‘if x spends money, this is disgusting’.<sup>23</sup> The same facial expression could also follow the Verb Phrase, in which case it arguably behaves as a post-speech gesture and plausibly triggers a supplement. Finally, as argued above, the iconic modulations of *GROW* are best compared to those of *looong*: at-issue contributions are made in both cases.

Three conclusions are worth highlighting. First, iconic enrichments (co-speech and post-speech gestures as well as iconic modulations) can be handled by established semantic mechanisms, although one case (cosuppositions) requires adjustments to presupposition theory (by introducing conditionalized presuppositions<sup>24</sup>). Second, there is no categorical difference between iconic enrichments in speech and in sign language: the same abstract typology is found in both modalities. Third, gestural enrichments do not make the same type of (at-issue) contribution as iconic modulations. But iconic modulations are arguably rich in sign language, and impoverished in speech. This yields systematic differences between sign with iconicity and speech with gestures.

<sup>23</sup> Importantly, this facial expression is not a grammatical marker, unlike other sign language facial expressions: it seems genuinely gestural.

<sup>24</sup> See Esipova (2019) for a slightly different formulation, on which co-speech gestures are presupposed to be redundant.

Importantly, the typology in (23) could be expected to apply to further types of enrichments, such as ‘vocal gestures’ (e.g. Schlenker, 2018b). Pasternak (2019) and Pasternak and Tieu (2022) argue that sounds that co-occur with speech (e.g. sound effects in radio drama) behave like co-speech gestures and yield an inferential profile characteristic of gestural cosuppositions. Perhaps surprisingly, we will see in Sect. 5.4 that music that co-occurs with films or cartoons might behave like co-speech gestures in triggering cosuppositions.

## 4.2 Iconic replacements in the typology of linguistic inferences: replicating the typology without words

The semantic difference between co-, post- and pro-speech gestures is certainly due to the *manner* in which they are realized, namely as co-occurring, following or replacing a word (see Schlenker, 2018b; Esipova, 2019 for possible explanations). But in addition, recent research suggests that pro-speech gestures alone neatly fill established categories of the inferential typology of language, which includes not just at-issue entailments and supplements, but also implicatures and (standard, non-cosuppositional) presuppositions, among others: depending on their informational content, they may make *additional* contributions that reflect inferential types (and probably algorithms) that are found in normal words.

These gestural findings were obtained with experimental means in Tieu et al. (2019). Moreover, Tieu et al. (2019) go one step further and replicate the typology in paradigms in which gestures are replaced with visual animations that the subjects could not have seen in a linguistic context before. We thus observe a high degree of productivity when it comes to dividing new semantic content ‘on the fly’ among established categories of the inferential typology. This argues for the existence of cognitive algorithms that make it possible to do so. For brevity, we discuss just two cases: iconic implicatures and iconic presuppositions.

### 4.2.1 Iconic implicatures

In some cases, the existence of inferences that arise ‘on the fly’ is expected by current theories. Consider the case of scalar implicatures. In (27), a gesture representing a partial wheel-turning is contrasted with a complete wheel-turning. It can be checked in separate examples that not *TURN-WHEEL* can mean ‘not turn the wheel at all’ (rather than ‘not turn the wheel exactly as depicted’, for instance). This suggests that the partial wheel-turning (i.e. *TURN-WHEEL*) can have a weak meaning, akin to ‘turn the wheel’. But as soon as this gesture evokes (thanks to the context) a more informative alternative *TURN-WHEEL-COMPLETELY*, an implicature is derived: John will *TURN-WHEEL* is understood to mean that John will turn the wheel but not completely. This was shown in experiments where participants endorse the target inference in (28)a significantly more than the control inference in (28)b.

(27) Context: John is training to be a stunt driver. Yesterday, at the first mile marker,

he was taught to *TURN-WHEEL-COMpletely* . Today, at the next

mile marker, he will *TURN-WHEEL* . (Tieu et al., 2019)

- (28) a. Target inference: John will turn the wheel, but not completely.  
 b. Baseline inference: John will turn the wheel completely.

Tieu et al. (2019) show in separate conditions that the gestures involved are unlikely to be mere codes for words because they have iconic implications that mere words would lack, for instance about the size of the wheel. In addition, similar results are obtained when gestures are replaced with artificial visual animations that subjects couldn't have seen before. As expected, then, implicature derivation is a fully productive process.

#### 4.2.2 Iconic presuppositions

In contrast with scalar implicatures, presuppositions are typically thought to be encoded in the lexical meaning of words (e.g. Heim, 1983), although there have been various attempts to propose 'triggering algorithms' that *deduce* the presupposition of an expression from its informational content (see for instance Abrusán, 2011 and Schlenker, 2021b for discussion). Strikingly, pro-speech gestures can trigger presuppositions, as can be illustrated by a modification of our *TURN-WHEEL* example: the question in (29) triggers the inference that Sally is behind the wheel, hence a significantly stronger endorsement of the inference in (30)a than in (30)b (Tieu et al., 2019). Many further tests have been used in the literature to argue that this and other gestural examples genuinely trigger presuppositions (e.g. Schlenker, 2019b).

(29) Jake and Lily are watching their four children ride bumper cars at the carnival. Each bumper car has two seats. As one of the bumper cars nears a bend in the track, the parents wonder:

Will Sally *TURN-WHEEL*  ?

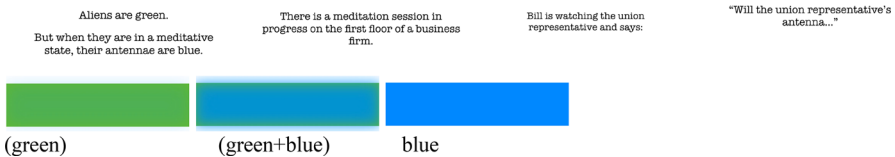
- (30) a. Target inference: Sally is in the driver's seat.  
 b. Baseline inference: Sally is in the passenger seat, not the driver's seat.

Going one step further, Tieu et al. (2019) carry out an experiment showing that the generation 'on the fly' of presuppositions from iconic material extends to stimuli

that subjects could not possibly have seen before. For instance, take a visual animation, described in (31), which represents an alien changing color from their normal state (green) to their meditative state (blue); for participants in the experiment, a change from green to blue triggered an inference that the aliens were not initially meditating, which is in line with how natural language expressions and gestures behave: in change of state constructions, the initial state is presupposed (e.g. Abrusán, 2011). This inference is preserved in questions, a characteristic property of presuppositions (illustrated in (12)a in Sect. 2.3.2). Further tests and further pictorial representations that involve changes of state were used to buttress the conclusion that visual animations too can trigger presuppositions.

**(31) Pictures from Tieu et al.’s videos testing presuppositions generated by visual animations**

(here: a change of state animation pertaining to an alien’s antenna turning from green to blue; original video: <https://youtu.be/U6dfs-XI2-4>)



**4.2.3 Going further**

The productive nature of implicature and presupposition generation extends to further inferential types. For instance, the contrasts illustrated in (26), pertaining to supplements, were obtained with gestures and visual animations in Tieu et al.’s experiment. Still further inferential types (e.g. pertaining to so-called ‘homogeneity inferences’) follow the same logic (Schlenker, 2019b; Tieu et al., 2019), suggesting that iconic content can be productively divided among a rich inferential typology. Migotti and Guerrini (2023) argue that the main results can be extended to auditory stimuli, specifically to pro- and post-speech musical excerpts, highlighting the generality of the phenomenon: the inferential typology can be replicated not just with gestures and visual animations but also with diverse non-linguistic sounds, when they are combined with natural language expressions.

A key question for future research is whether this productive division of semantic content among established slots of the inferential typology only arises when the stimuli are embedded in sentences, or whether some of them might arise in visual narrations that are not linguistically embedded (and are thus used in a more standard fashion)—a tantalizing if remote possibility.

## 5 Linguistics beyond words II: visual and musical narratives

Strikingly, recent research has argued that methods inspired by formal linguistics can illuminate three kinds of non-linguistic objects: visual narratives, music and dance.

### 5.1 Visual narratives

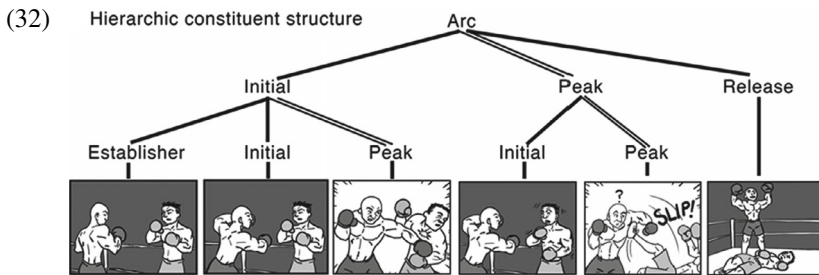
#### 5.1.1 Syntax of visual narratives

Starting with visual narratives, such as comics, recent literature (Cohn, 2012, 2018, 2020) argues that an independent syntax of visual narratives is required in addition to a semantic analysis (see Sect. 5.1.2), though the two share an interface with each other.<sup>25</sup> Initial studies, discussed in Cohn et al. (2012), argue for the existence of syntactic structure in visual sequences by replicating the core point behind Chomsky's (1957) *Colorless green ideas sleep furiously*, demonstrating that syntactic well-formedness is independent from semantic well-formedness. The experiments compared (i) comics that told semantically coherent narratives to (ii) comics that followed narrative structuring principles (i.e. syntax) but lacked semantic connection between the panels, as well as (iii) scrambled comics (which lacked narrative structure, thus lacking both syntax and a semantic connection). While the latter two (ii+iii) gave rise to increased N400 brain responses, indicating a problem with semantic processing, the meaningless narratives that obeyed structuring principles (ii) were closer to the coherent narratives (i) in response times. This conclusion is supported by further ERP measures; apart from the lack/presence of an N400 effect, the authors also measured the presence of LANs (left anterior negativity) and P600 effects that arise from manipulations of the narrative grammar; they find that these effects mirror ones that arise from manipulations of natural language syntax, and conclude that narrative grammar is distinct from semantics.

---

<sup>25</sup> The work discussed in this section is part of a more extensive scholarly exchange. In response to earlier syntactic proposals (e.g. Cohn, 2012; Cohn et al., 2012), Bateman and Wildfeuer (2014) propose a syntax-free semantic approach; their syntax-free approach, in turn, is critiqued by Cohn (2018), who defends a syntactic approach based on empirical evidence. The issue in that debate is that it seems self-evident that visual narratives must have a semantics, but less self-evident that it has a syntax. We discuss the semantics of visual narratives in Sect. 5.1.2.

As discussed in Sect. 2.1.2, a notion of constituency and grouping is at the heart of syntactic analysis. Core tenets of a syntactic approach to visual narratives include higher-level and lower-level grouping, much in line with the syntactic analysis of other non-standard objects, such as music and dance, as we will see in Sects. 5.2.1 and 5.3.1. For visual narratives, Cohn proposes that their structure consists of narrative categories (e.g. *Establisher*, *Initial*, *Peak* and *Release*), grouping (constituency), and hierarchical structure, similar to those present in natural language<sup>26</sup>. An application of syntactic analysis to comics is illustrated by the example in (32),<sup>27</sup> a comic strip depicting a boxing match, where one of the fighters slips (presumably on purpose) and forfeits the match, after having been struck just once. The depicted grouping structure reflects the ebb and flow of tension in the narrative (Initial-Peak-Release), modeled by the assignment of narrative categories to the panels; the four main narrative categories are defined in (33).



Originally Figure 4(b) from Cohn (2020: 361), © Neil Cohn

Published under Attribution-NonCommercial-NoDerivatives 4.0 International licence (CC BY-NC-ND 4.0, <https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Each narrative category is determined on the basis of diagnostics (see Cohn, 2015 for details). Although the syntax of visual narratives employs grouping and constituency, and the narrative categories are in and of themselves not considered to be semantically determined, they are taken to interface with discourse semantic properties of the panels. (Cohn assumes a relationship between grammar/syntax and semantics that is conceived of as an expansion of Jackendoff's (2002) *Parallel Architecture*.)

(33) *Basic narrative categories* (quoted from Cohn, 2018: 2–3)<sup>28</sup>

- a. Establisher (E) – sets up an interaction without acting upon it, often as a passive state.
- b. Initial (I) – initiates the tension of the narrative arc, prototypically a preparatory action and/or a source of a path.

<sup>26</sup> An open question that arises at this point is whether these structural categories apply to stories in general rather than specifically to comics.

<sup>27</sup> The arrangement of panels in actual comics varies; for the purposes of analysis, the panels are given in a linear (left-to-right) fashion; see Cohn (2018) for discussion.

<sup>28</sup> See Cohn (2018) for further refinements of the system.

- c. Peak (P) – marks the height of narrative tension and point of maximal event structure, prototypically a completed action and/or goal of a path, but also often an interrupted action.
- d. Release (R) – releases the tension of the interaction, prototypically the coda or aftermath of an action.

While grouping structure is assumed to arise in the mind of a comic reader, a researcher can utilize diagnostics to generate and analyze it: the first step involves identifying the Peak(s) in the visual sequence, from which a researcher works backwards to identify the Initial and Establisher (building up to the Peak) and finally the Release<sup>29</sup>. Higher/intermediate level grouping applies the same narrative categories in a recursive manner; in the example above, the first Peak is less prominent than the second one, where the narrative tension culminates. This correlates with the observation that the narrative is disrupted more when the second Peak is deleted than when the first Peak is deleted.<sup>30</sup> Therefore, at the higher level of structure, the constituent containing the second Peak functions as the (higher-level) Peak, whereas the constituent containing the first Peak becomes the Initial.<sup>31</sup> Crucially, if deletion of the first Peak in a similar example had a bigger effect than deletion of the second Peak, then different narrative categories would emerge at the higher level: the constituent containing the first Peak would function as the higher-level Peak, while the constituent containing the second Peak would be part of the Release.

### 5.1.2 Semantics of visual narratives

In connection with the introduction of iconic semantics, Sect. 3.1 also introduced a projection-based semantic of pictures; a gulf separates the semantics of individual pictures from the semantics of visual narratives, spanning from comics to films. How can formal semantic analysis be applied to comic strips such as the example in Sect. 5.1.1?

Quite generally, formal pictorial semantics is in its infancy, but important progress is being made on several fronts. First, pioneering work on the visual language of comics is getting the recognition it deserves, from sophisticated analyses of visual morphology (e.g. reduplication used to evoke movement, plausibly drawing on the same iconic resources discussed in Sect. 3) to a theory of narrative structure in comics, as discussed above (Cohn, 2013). Second, Cumming et al. (2017) investigate in formal detail constraints on viewpoint shift in film, i.e. changes of camera angle that are permissible in view of established narrative conventions. Third, Abusch and Rooth (2017) have upgraded Greenberg's projection-based semantics to analyze visual narratives. The simplest example

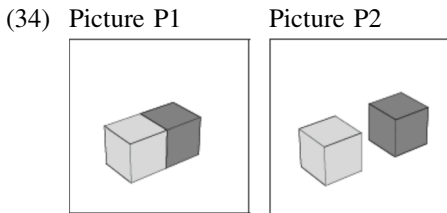
<sup>29</sup> This is quite reminiscent of syllable structure, Onset-Nucleus-Coda, in phonology. See also Jackendoff (2009) for a broader discussion of the structure of complex actions in general.

<sup>30</sup> Deletion is one of the diagnostics used to identify narrative categories, see Cohn (2015).

<sup>31</sup> The top-level (root) node *Arc* contains the entire narrative (story arc), and serves no further narrative role.



appears in (34), which is a minimalistic picture sequence that contains two differently colored cubes of the type illustrated in example (15) of Sect. 3.1. When we view this picture sequence, we infer that the light grey cube in Picture P2 is the same as the light grey cube in Picture P1, and the same for the dark grey cube; moreover, we infer that the distance between them is somehow increasing in that they start out by touching one another in P1, which is no longer the case in P2 (Abusch and Rooth, 2017).



The fact that the two pictures are arranged as a narrative sequence provides information about the situations described, as well as their ordering. Recall our definition of truth (and thus of “true”) in pictures, from Sect. 3.1, where a given picture  $P$  is true in a situation  $s$  if the situation  $s$  can project onto  $P$  in relation to a viewpoint  $v$ . In essence, we can generalize in such a way that a series of  $n$  pictures  $\langle P_1, \dots, P_n \rangle$  is true of  $n$  situations  $\langle s_1, \dots, s_n \rangle$  just in case the  $s_i$ 's are temporally ordered in the right way, and each  $s_i$  projects onto the corresponding  $P_i$  relative to the same viewpoint  $v$ ,<sup>32</sup> as stated in (35), which is a temporal version of (16).

(35) **Picture sequences true of  $n$  situations** (after Abusch, 2013, 2020)

A picture sequence  $\langle P_1, \dots, P_n \rangle$  is true of situations  $\langle s_1, \dots, s_n \rangle$  **relative to viewpoint  $v$**  along the system of projection  $S$  just in case:

- (1) temporally,  $s_1 < \dots < s_n$ ;
- (2)  $\text{proj}_S(s_1, v) = P_1$  and ... and  $\text{proj}_S(s_n, v) = P_n$ .

While the situation-based analysis of picture sequences in (35) is a powerful tool to study visual narratives, Abusch convincingly argues that this framework would yet be insufficient, in particular because narrative sequences give rise to ambiguities of cross-reference. Concretely: the second picture of (34) is most naturally interpreted as involving the same cubes as the first, but nothing fully excludes the possibility that the dark cube disappeared and was replaced by another dark cube a bit further away. A semantics that is based exclusively on temporal sequences of pictorial situation-descriptions, as in (35), doesn't suffice to derive the most salient

<sup>32</sup> For now, we simplify by glossing over viewpoint changes, limiting ourselves to a single viewpoint  $v$ . Similarly, we do not consider changes in the projection system  $S$  within a picture sequence. Picture sequences that involve a non-linear temporal development will also require a more complex semantic analysis; this includes visual sequences that ‘rewind time’, incorporate flashbacks, or involve dream sequences.

reading (same cubes in both pictures), nor the ambiguity that permits less plausible readings (where one or both cubes have been replaced by identical cubes). Abusch (2013) suggests that the salient interpretation is obtained because visual representations contain abstract variables (which she relates to Pylyshyn's (2003) indexes in vision). A projection-based semantics can thus be combined with variables to derive the meaning of visual narratives (see Abusch and Rooth (2017) for further pictorial operations possibly reminiscent of language; in terms of the semantic model, Abusch (2013) implements her approach in Discourse Representation Theory, an approach to visual narratives that is shared with Bateman and Wildfeuer (2014), Maier and Bimpikou (2019), and Maier (2019)).

## 5.2 Musical narratives

While it usually steers clear of a direct importation of linguistic theories, music cognition has been inspired by linguistic methods in two areas: the study of music syntax and the more recent exploration of music semantics.

### 5.2.1 Basic music syntax

The analysis of Western tonal music has led to the belief that music has a sophisticated syntax. Upon venturing into this domain, researchers find themselves with an embarrassment of plenty: different authors posit different syntactic systems, sometimes with different goals.

In their pioneering study of music syntax, Lerdahl and Jackendoff (1983), posit four levels of structure. As summarized in Lerdahl (2001a), their theory proposes:

*four types of hierarchical structure* simultaneously associated with a musical surface. *Grouping structure* describes the listener's segmentation of the music into units such as motives, phrases, and sections. *Metrical structure* assigns a hierarchy of strong and weak beats. *Time-span reduction*, the primary link between rhythm and pitch, establishes the relative structural importance of events within the rhythmic units of a piece. *Prolongational reduction* develops a second hierarchy of events in terms of perceived patterns of tension and relaxation. (Lerdahl, 2001a: 3; emphasis added)

To illustrate, (36) provides the metrical structure (the square horizontal brackets beneath the musical score) and the grouping structure (the round horizontal brackets beneath the musical score) of the beginning of Mozart's Sonata No. 11, K. 331.<sup>33</sup> To begin with, a reader should focus on the round horizontal brackets (the grouping structure), which we discuss first.




<sup>33</sup> As Lerdahl and Jackendoff (1983: 63) point out, an alternative grouping is conceivable, with the last E of the 1st bar grouped with the 2nd bar, and with the last D of the 2nd bar grouped with the 3rd bar.

- (36) Metrical structure [square brackets] and grouping structure [round brackets] of the beginning of Mozart’s K. 331 Sonata No. 11 (Lerdahl and Jackendoff, 1983)

[https://imslp.hk/linkhandler.php?path=/imglnks/euimg/0/09/IMSLP707462-PMLP1846-Piano\\_sonata,11,\\_in\\_A\\_major.mp3](https://imslp.hk/linkhandler.php?path=/imglnks/euimg/0/09/IMSLP707462-PMLP1846-Piano_sonata,11,_in_A_major.mp3)



As illustrated by the round brackets in (36), grouping yields a hierarchical, tree-like structure (the bottom-most round bracket including all other brackets, with two round brackets in the next line above it, which corresponds to a lower-level grouping). Such **grouping structure** is “an auditory analog of the partitioning of the visual field into objects, parts of objects, and parts of parts of objects” (Lerdahl and Jackendoff, 1983: 36). Principles of grouping thus derive from Gestalt principles of perception (Wertheimer, 1923), with local preference rules that specify, for instance, that group boundaries are preferably created by (i) large pitch intervals, and (ii) pauses.<sup>34</sup> To illustrate (i), the 3rd note in (37)a,b is preferably grouped with the first two notes because there is a large pitch interval between the 3rd and 4th note. Contrastively, the 3rd note is preferably grouped with the last two notes in (37)c,d, because there is a large pitch interval between the 2nd and the 3rd note.

- (37) a.  b.  c.  d. 

To illustrate (ii), the pauses (rests, ♯ and 7) in (38)a-c create group boundaries, and in fact override the preference that might be expected from (i): in (38)b,c, despite the large pitch interval between the second and the third note, they are grouped together due to the pause that separates the first three notes from the last two.

<sup>34</sup> Further principles help determine group boundaries at higher levels, e.g. pertaining to the optimal size of groups, or to optimal parallelism among groups that look similar.



- (40) Time-span reduction obtained from (36) by selecting in each the musical event which is metrically strongest/harmonically most stable (Lerdahl and Jackendoff, 1983)

The image shows a musical score in G major, 3/4 time, with a treble and bass clef. The melody is in the treble clef, and the accompaniment is in the bass clef. Below the score, there are four levels of brackets representing time-span reductions. The first level shows individual chords: I, I<sup>6</sup>, V<sup>6</sup>, V<sup>4</sup>/<sub>3</sub>, "vi<sup>7</sup>", V<sup>6</sup>, I, V. The second level groups I and I<sup>6</sup> together, V<sup>6</sup> and "vi<sup>7</sup>" together, and I and V together. The third level groups I, I<sup>6</sup>, V<sup>6</sup>, and "vi<sup>7</sup>" together, and I and V together. The fourth level groups the entire first four chords (I, I<sup>6</sup>, V<sup>6</sup>, V<sup>4</sup>/<sub>3</sub>) together, and the last four chords (I, V, "vi<sup>7</sup>", V<sup>6</sup>) together.

The outcomes of calculating a time-span reduction are headed trees that are somewhat reminiscent of natural-language syntax, but obtained from entirely different considerations (e.g. unlike syntactic constituents in language, grouping structure in music is derived from principles of general perception, not by phrase-structure rules<sup>37</sup>). Lerdahl and Jackendoff (1983) further posit yet another structure, called **prolongational reduction**, whose function is to provide a hierarchy of events “in terms of perceived patterns of tension and relaxation” (Lerdahl, 2001a), a specifically harmonic notion. Prolongational reduction is the hierarchical structure, out of the four that were introduced for music, that is most strongly tied to music—it might be non-trivial to find it in visual narratives (Sect. 5.1) or in dance (Sect. 5.3), where we do find counterparts of grouping structure.

The importance of prolongational reduction is evident in more recent research on music syntax, where the analysis of ‘harmonic syntax’, which roughly corresponds to prolongational reductions, has given rise to active debates. While Lerdahl and Jackendoff (1983) stated all their structural rules as parsing rules (with preference principles guiding the listener towards preferred structures), Pesetsky and Katz (2009) propose to define the same system in terms of generation, highlighting the similarity with linguistic syntax. Departing even further from Lerdahl and Jackendoff’s framework, Rohrmeier (2011) proposes phrase-structure rules to account for certain aspects of harmonic syntax in Western tonal music, while Granroth-Wilding and Steedman (2014) propose a combinatorial categorial grammar (and parser) for jazz harmonic progressions. The latter is even associated with a semantics, one that encodes motion in tonal pitch space.

Strikingly, all these frameworks propose syntactic formalisms that go beyond the finite-state bound, and some (Rohrmeier, 2011) have developed precise formal arguments to the effect that this is not just a matter of convenience: some properties of (Western tonal) music necessitate recursive mechanisms that are not finite-state. While the debate is still to some extent open (see Rohrmeier et al., 2015 for

<sup>37</sup> Katz (2022b) has emphasized similarities between grouping structure in music and prosodic grouping in language (rather than syntactic constituency).

discussion),<sup>38</sup> the overall picture is that music can probably not be accounted for by purely finite-state methods. However for most authors (and with the notable exception of Pesetksy and Katz, 2009), the syntax of music is very different from the syntax of human language, although it might share some modules with human prosody (for discussion, see Jackendoff, 2009; Lerdaahl, 2001b, 2013; Katz, 2022a, 2022b).

### 5.2.2 Basic music semantics

While it is relatively uncontroversial that music has a sophisticated syntax, the existence of a music semantics is considerably more debated. By “music semantics”, we will mean here a semantics in the usual sense, i.e. one that provides information about the extra-musical world.<sup>39</sup> But is there such a thing?

To see that pitch alone can convey meaning, consider a sound-based variant of the fuel level gauge that we briefly discussed at the beginning of Sect. 3. Imagine a scenario where a pilot pushes a button and hears a single tone in response; imagine that the pitch of the tone strictly corresponds to the amount of fuel in the tank (the higher the pitch, the more fuel in the tank); this would be a simple case of iconic semantics in the sense of Greenberg (2021b).

Of course, music is much more complex than this toy example, but related (albeit more abstract) iconic inferences are broadly attested. There are numerous experimental and introspective investigations of inferences triggered by actual musical pieces about the extra-musical world. An initial list of systematic effects appears in (41); this list makes reference to an object (sometimes called a ‘virtual source’, or ‘denoted object’) that is described by the music. To illustrate (41)a(i), the French composer Saint-Saëns famously used a double bass to evoke an elephant, the virtual source/denoted object, in his *Carnival of the Animals*. If a MIDI rendition is raised by three octaves so as to be high-pitched rather than low-pitched, the inference is reversed: instead of a large object, a small one is evoked (original: <http://bit.ly/2mea8pQ>; three octaves higher: <http://bit.ly/2CI6Xhk>). Potential denotations are diverse: they are not limited to movable objects such as animals, but could be landscapes described by the music, or even (in refinements of the analysis) emotions.

<sup>38</sup> Beyond Western tonal music, Rohrmeier et al. (2015) write: “There are also analytical findings that suggest that principles of hierarchical organization may be found in classical North Indian music that is based on a tradition of extensive oral teaching. However, more cross-cultural research on other cultures and structure in more informal and improvised music is required [...]”

<sup>39</sup> The terminology can occasionally be confusing: for several authors, musical meaning has to do with inferences about certain properties of the musical *form*, rather than about a music-external reality. For instance, Meyer (1956) writes that “one musical event (...) has meaning because it points to and makes us expect another musical event”; this gives rise to expectations and emotions that constitute what Meyer calls “embodied meaning”. Huron (2006) argues that various emotions of a musical or extra-musical nature derive from general properties of expectation, or in other words of our attempts to anticipate what will come next, in music or elsewhere: the inferences are thus about musical form, although they have emotional consequences. As for the semantics of Granroth-Wilding and Steedman (2014), it pertains to movement in tonal pitch space, not to the extra-musical world.

- (41) Examples of inferential effects (Schlenker, 2018d: Appendix II, with links to examples)
- a. Lower pitch may indicate that the denoted object (i) is larger, or (ii) is less excited/energetic.
  - b. Lower loudness may indicate that the denoted object is (i) less energetic, or (ii) further away.
  - c. Lower speed may indicate that the denoted object is slower.
  - d. Silence may indicate that an event is interrupted.
  - e. Lesser harmonic stability may indicate that the denoted is in a less stable (i) physical or (ii) emotional position.
  - f. A change of key may indicate that the denoted object is moving to a new environment.

How should these inferences from music be explained?<sup>40</sup> In auditory or visual perception, agents seek to find information about the causal sources of their percepts. In auditory perception, certain sounds reach the human ear and, depending on their properties, give rise to information about the surrounding objects and events: one may hear a low-frequency call produced by an animal and infer that it is very large; or one may hear a car engine whose loudness decreases and infer that it is moving away. It has been suggested (e.g. Schlenker, 2017a, 2019a, 2022) that the same general idea applies to music semantics: the listener seeks to draw inferences about certain objects (e.g. an imagined animal in Saint-Saëns' narrative) based on some properties of the musical sounds (such as low pitch). These acoustic properties usually trigger certain inferences about their causal source (for instance: low pitched sounds tend to be produced by large objects and animals), and these inferences are then applied to the fictional objects the music is 'about'.

Inferential rules of the type in (41) have been argued to be of two kinds. Some are derived from normal auditory cognition, as in the case of a low-frequency sound used to evoke a large object: the inferential rule in (41)a(i) presumably owes to the fact that large resonance chambers produce lower-pitched sounds. Similarly, all other things being equal, a sound is perceived as softer when its source is further away, explaining the inferential rule in (41)b(ii).<sup>41</sup> But other inferences are more specifically musical in nature, particularly when it comes to harmonic notions. A key notion in music is that of consonance vs. dissonance, and it can produce powerful inferential effects. To illustrate, Saint-Saëns uses a radically slowed down version of the French *Can Can* dance to evoke tortoises (see <https://youtu.be/6HQqaKEz4tg>). Later in the piece, dissonances were suggestive of the tortoises tripping (see the simplified version in <https://youtu.be/UqUQQORfCMY><sup>42</sup>). (41)e(i) makes use of the distinct but related notion of harmonic stability, which is relative to

<sup>40</sup> The discussion in this paragraph (including some of the wording) follows Schlenker (2022).

<sup>41</sup> In the area of inferences derived from normal auditory cognition, a particularly fruitful line of research has focused on properties of animal and human vocal signals that can be used in music to trigger powerful emotional effects (e.g. Aucouturier et al., 2016; Liuni et al., 2020).

<sup>42</sup> Rewriting the music so as to remove the dissonances removes the inference as well (see the simplified version in [https://youtu.be/0A2egp\\_OIVU](https://youtu.be/0A2egp_OIVU), rewriting by A. Bonetto).

a key and thus to a musical context.<sup>43</sup> Such harmonic inferences do not have counterparts in normal auditory cognition, and semantic inferences from harmonic properties of the music may even be culture-dependent, specific to a given musical tradition.<sup>44</sup>

If music indeed triggers inferences about the extra-musical world, can we define the semantic content of a musical excerpt? This was done in a simplified framework by positing that a series of  $n$  musical events is true of an object undergoing  $n$  events in the (real or imagined) extra-musical world just in case the inferences triggered by the music are all satisfied by the corresponding events. Here it matters that all the inferential effects in (41) are of the following form: *If musical events  $M_1$  and  $M_2$  stand in relation  $R$ , their respective denotations  $s_1$  and  $s_2$  stand in relation  $R^*$* . (For example, in (42) below, one rule has  $R =$  ‘is less loud than’ and  $R^* =$  ‘has less energy than’.) A semantics can for this reason be defined by requiring that certain relations among musical properties be preserved by the events depicted by the music. This makes it possible to define a notion of truth for musical sequences, partly similar to the notion of truth for pictorial sequences in (35) (here we use the term ‘eventualities’ rather than ‘situations’ because it is more intuitive, but the structure of the account is similar):

(42) **Musical sequences true of  $n$  events** (after Schlenker, 2017a, 2019a)

A musical sequence  $\langle M_1, \dots, M_n \rangle$  is true of an object undergoing eventualities  $\langle e_1, \dots, e_n \rangle$  relative to auditory point  $v$  just in case:

- (1) temporally,  $e_1 < \dots < e_n$ ;
- (2) certain preservation conditions are satisfied, for instance ones corresponding to (41)b,e:
  - a. If  $M_i$  is less loud than  $M_k$ , then (i)  $e_i$  has less energy than  $e_k$ ; or (ii)  $e_i$  is further from the auditory point  $v$  than  $e_k$  is.
  - b. If  $M_i$  is less harmonically stable than  $M_k$ , then  $e_i$  is less
    - (i) physically, or (ii) emotionally stable than  $e_k$ .

In this simple proof of concept, it can already be seen that the inferences triggered will be very general, hence the abstract character of musical meaning: an excerpt will typically be true of lots of very diverse tuples of events (namely those that satisfy the preservation conditions). It can be highlighted by revisiting Leonard Bernstein’s celebrated discussion of Richard Strauss’s *Don Quixote (Variation II)* (link: <https://youtu.be/dbGV-gUsEPI>). As an instance of ‘program music’, this piece has an explicitly descriptive character. Yet Bernstein sought to convince his audience that *even* such a piece does not describe any real or imagined situation: instead,

<sup>43</sup> For instance, in the key of C major (corresponding to the scale obtained by pressing the white keys on a piano, starting with C), C is the most stable note, and the C major chord (= CEG) is the most stable chord.

<sup>44</sup> For instance, McAuley et al. (2021: 2) observe that an atonal piece of music triggered very different inferences in a cross-cultural study; participants in Arkansas and Michigan associated the atonality with horror and murder, while participants in Dimen, China interpreted the atonality very differently, “involving happy games with friends”. These findings do not address the question of whether the harmonic inferences in (41) are themselves culture-specific; happy games may well be conceptualized as involving physical instability.



he argues that the true meaning of music is “the way it makes you feel when you hear it”. To make his point, Bernstein set out to tell the *wrong* story, and showed that it could be made to fit the music just as well as the intended story. The original one is about Don Quixote (i) departing on his horse to conquer the world, (ii) approaching a flock of sheep baaing, which he mistakes for an army, (iii) charging at them and creating chaos in the process, and finally (iv) feeling proud about his knightly deed. Bernstein’s *wrong* story pertained to Superman (i’) departing on his motorcycle to free an unjustly imprisoned friend of his, (ii’) approaching the jail, in which the prisoners are snoring, (iii’) charging into the prison and wreaking havoc in the process, and finally (iv’) triumphantly hurling his friend back to freedom. It is striking that Bernstein’s two stories are, with minor exceptions, isomorphic to each other: there is a close correspondence to the events described in (i)-(i’), (ii)-(ii’), (iii)-(iii’) and (iv)-(iv’).

The correspondence is not due to lack of imagination on Bernstein’s part, but to the music itself. Specifically, the details of the music trigger inferences that must be preserved by both stories—which is unsurprising if music has a semantics (Schlenker, 2022). For instance, the triumphant character of the beginning, (i)-(i’), is in part due to the upwards melodic movement, as heard in (43)a; importantly, when the music is rewritten in accordance with composition rules so as to invert the melodic movement, (43)b, this triumphant character disappears, as expected by (41)a(ii). The sheep baaing, (ii), and the prisoners snoring, (ii’), create a somewhat chaotic effect in the story, evoked by dissonances in the music, as heard in (44)a; again, when the dissonant chords are replaced with consonant ones, (44)b, the chaotic effect largely disappears, as expected by (41)e(i). Similarly, the sheep and prisoners appear to be approaching the perspectival center, (iii)-(iii’), because of a crescendo (increasing loudness) in the music, as heard in (45)a; once again, when the crescendo is replaced with a diminuendo (decreasing loudness), (45)b, the effect disappears (if anything, the denoted objects appear to move away), as expected by (41)b(ii).

- (43) a. Original: upwards melodic movement at the beginning (simplified MIDI) [https://youtu.be/\\_dSwjTMyzSM](https://youtu.be/_dSwjTMyzSM)  
 b. Rewritten: downwards melodic movement at the beginning (A. Bonetto) <https://youtu.be/5e-39sxEKhk>
- (44) a. Original: dissonant chords (simplified MIDI) <https://youtu.be/fKgJDy0wYk0>  
 b. Rewritten: consonant chords (A. Bonetto) <https://youtu.be/EnhSaeMORCK>
- (45) a. Original: crescendo (simplified MIDI) [https://youtu.be/\\_mMA9dByPAw](https://youtu.be/_mMA9dByPAw)  
 b. Rewritten: diminuendo (A. Bonetto) <https://youtu.be/kCfay6s4Igs>

In sum, one’s ability to tell a ‘wrong’ story to fit the music doesn’t show that the latter has no semantics, just that it has a fairly abstract one, i.e. that its meaning singles out rather diverse situations.

One key question for future research pertains to the interface between music syntax and music semantics. It has been suggested (rather speculatively) that

Lerdahl and Jackendoff's grouping structure arises from an attempt to recover the structure of the events denoted by the music (Schlenker, 2017a, 2019a; Zaradzki, 2021). In essence, the idea was that events come with a natural part-of structure, and that when further conditions are imposed, a tree structure can be derived, but with some occasional exceptions that Lerdahl and Jackendoff noted.<sup>45</sup> Similarly, the headed nature of groups might be reanalyzed in semantic terms, with some events being construed as more central than others. Whether this semantic reanalysis of grouping structure and time-span reductions is correct remains to be seen; and the interface between further music structures and music semantics has yet to be investigated.<sup>46</sup>

### 5.2.3 Adding variables

The simple cases in Sect. 5.2.2 involved excerpts with a single musical voice corresponding to a single denoted object; in more complex cases, a listener may infer that there are several denoted objects, which raises a question analogous to the one Abusch (2013) asked about visual narratives: what are the cross-reference relations among them? It may thus be fruitful to enrich musical representations with variables, just as was argued for visual narratives by Abusch (here we follow Schlenker, 2022).

This issue is best illustrated by an example. The beginning of Chopin's Mazurka op. 33 no. 2 has an extremely simple structure that involves two melodies, A followed by B, a sequence which is then repeated. We can write the form of the piece as AB A'B' where A' and B' are the repetitions of A and B; that is, A' = A and B' = B. In the original score, (46)b, the first [AB] sequence is played loudly (*forte/f*), whereas the repetition [A'B'] is played softly (*pianissimo/pp*). Crucially, this affects a listener's inferences on the nature of the denoted objects. If we manipulate the music so that it is given a flat realization with constant loudness, as in (46)a, it is compatible with multiple interpretations: listeners can infer (i) a single object that is evoked by the entire piece, or, alternatively, (ii) one object that is associated with [AB] and a second object that is associated with [A'B']; possibly, a reader may even infer (iii) one object that is associated with A and A', and another one that is associated with B and B'. This ambiguity of (46)a, in effect, is an ambiguity in cross-reference, just as was the case with Abusch's cube in (34), or for that matter with (9)a in Sect. 2.2.2.

<sup>45</sup> A particularly interesting case arises in structures of the form  $[a [b] c]$ , where b belongs both to the group [a b] (= lower brackets) and [b c] (= upper brackets). Lerdahl and Jackendoff (1983) argue that such cases arise in grouping in visual perception as well, for instance in case an object is naturally construed as being a part of two larger, overlapping objects. This is reinterpreted in Schlenker (2017a, 2019a) as a case in which, in the denotations of the music, a subevent is naturally construed as being part of two temporally overlapping events (neither of which is a subpart of the other). See Migotti and Zaradzki (2019) and Zaradzki (2021) for more subtle conditions on the syntax/semantics interface in music.

<sup>46</sup> Lerdahl (2022) sketches a theory of musical narratives that is rooted in patterns of tension and relaxation, and thus in prolongational reductions. Lerdahl's analysis does not presuppose the existence of a music semantics. We leave open what relation, if any, Lerdahl's analysis (and more generally music syntax) might bear to Cohn's syntactic structures for comics (see Sect. 5.1.1 above).

- (46) a. Flat realization [AB] [A'B'] [https://youtu.be/IMZrU\\_vEA7A](https://youtu.be/IMZrU_vEA7A)  
 b. Chopin's dynamics [AB]<sub>f</sub> [A'B']<sub>pp</sub> <https://youtu.be/tyJ0ZQxG8rU>  
 c. Anti-Chopin dynamics A<sub>f</sub>B<sub>pp</sub> A'<sub>f</sub>B'<sub>pp</sub> <https://youtu.be/dYTIAdI4-B0>

We can now manipulate the piece in two ways that can help disambiguate. When we add the dynamics of Chopin's original score<sup>47</sup> in (46)b, (iii) appears to become far less likely. The reason is that [AB] is played *forte*, while [A'B'] is played *pianissimo*. This can be made sense of in an interpretation such as (ii), in which the first denoted object is energetic or close-by and corresponds to [AB], while the second is less energetic or further away and corresponds to [A'B']. Importantly, one can also create an 'anti-Chopin' dynamics, (46)c, in which A and A' are realized forte, while B and B' are realized pianissimo: this may then suggest reading (iii), where A and A' correspond to one source, while B and B' correspond to another, as can be heard in (46)c.

In several orchestrations written for a ballet by Michel Fokine, the identity of the denoted objects is made more salient by the use of different timbres, associated with different instruments: Chopin's dynamics [AB]<sub>f</sub> [A'B']<sub>pp</sub> (*forte*, then *pianissimo*) is reflected by the use of different timbres for [AB] and [A'B'], typically using the orchestra for [AB] and individual wind instruments for [A'B']. An example is Britten's orchestration in (47) (see Schlenker, 2022 for further examples of orchestration of the same piece that make related choices).

- (47) Benjamin Britten's orchestration (1941)<sup>48</sup>: [AB]<sub>orchestra</sub> [A'B']<sub>oboe+flute</sub> <https://youtu.be/aHUeA3526DY>

Finally, the importance of cross-reference becomes even more salient when a ballet co-occurs with the music. In Fokine's piece, [AB] corresponds to a movement of the main ballerina, [A'B'] to that of the other dancers.

- (48) Fokine's *Les Sylphides* (originally called *Chopiniana*), movement on Chopin's Mazurka op. 33 no. 2  
 Performance from 1984, American Ballet Theatre, on an orchestration close to Britten's version  
<https://youtu.be/Lsuc3KUKKpQ>  
 [AB]<sub>main ballerina</sub> [A'B']<sub>other dancers</sub>

<sup>47</sup> Chopin's score can be found at: <https://drive.google.com/file/d/15Vz7IKgRDKWawUTxW9ZKWs moaDT912A8/view?usp=sharing>.

<sup>48</sup> Retrieved from <https://www.youtube.com/watch?v=DQLde6QJXvM> on December 13, 2019. Credits: Chopin: *Les Sylphides* (Arr. for Orchestra by Benjamin Britten, Mono Version) Joseph Levine, American Ballet Theatre Orchestra, 1 January 1954.

In sum, an abstract notion of iconic semantics may be used to develop a semantics for music, and it too might benefit from the use of variables to enrich iconic representations, just as was argued by Abusch for visual narratives.

### 5.3 Dance

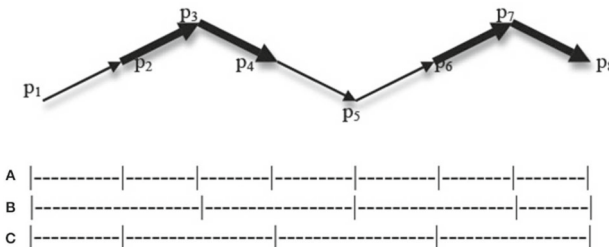
In Sect. 5.1, we discussed the syntax and semantics of visual narratives. In a further development of Super Linguistics, dance can be analyzed, at the very least, as a (3D) visual animation, but one with several peculiarities that require a special treatment: (i) it is produced by dancers, with physical and genre-specific constraints on permissible movements; (ii) depending on the genre, it may have a particularly abstract semantics (e.g. it need not be obvious what a ballet represents); (iii) it often stands in a particularly close relation to music.

#### 5.3.1 Dance syntax

If one asks whether dance may have a syntax (and thus potentially a ‘grammar’), two reasonable points of comparison are the syntax of visual narratives (Sect. 5.1.1), and the syntax of music (Sect. 5.2.1). Building on the latter, Charnavel (2019) proposes that dance is naturally perceived as hierarchically structured in ways that are reminiscent of musical grouping. Specifically, she argues with experimental means that some grouping preference rules proposed by Lerdahl and Jackendoff (1983) for music have analogues in the grammar of dance. A central insight may thus be that some version of grouping is a core concept that unifies all of the human modes of expression discussed so far: visual narratives, music, dance—and language, as discussed in Sect. 2.1.2.

Focusing on probing grouping preference rules in dance, Charnavel (2019) identifies six properties of dance that result in the formation of a group boundary: *change of {body part, orientation, level, direction, speed and quality}*. Concretely, (49) illustrates three alternative ways of generating grouping structure based on the movement of a dancer (schematically seen from above in the shape of an arrow).

(49) Three possible grouping structures for a single dance sequence (cited from Charnavel, 2019: Figure 5)

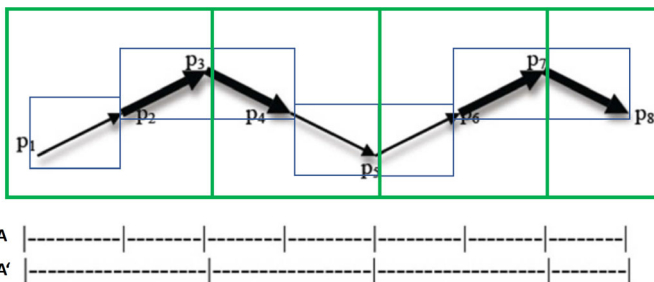


Each  $p_i$  indicates a potential grouping boundary. The arrows indicate the path of the dancer, whereas the thickness of the arrows reflects the speed of the movement,

for instance if the thicker arrow reflects a faster speed, the thinner arrow reflects a slower speed. Charnavel then considers three potential rules (labeled as A, B and C) according to which the movement could be sequenced. Rule A specifies that change of speed and change of direction each trigger a grouping boundary. By contrast, for Rule B, a single property, change of direction, triggers the appearance of a group boundary. For Rule C, a group boundary is created by change of speed and nothing else. To decide which rule is actually applied by onlookers, a partial answer can be found in an experiment that Charnavel (2019) carried out. Participants had to split a video into two sequences, and each stimulus video contained two out of the six possible changes that could mark a grouping boundary. The results give some indication as to whether participants would choose Rule B or Rule C, though they do not bear on whether participants would consider Rule A. Charnavel’s findings were that some properties are weighted more than others in determining group boundaries: change of body part had a stronger effect than all other changes—orientation, speed, etc. For the properties in our illustration, there was no statistically significant preference for either change of direction (B) or change of speed (C); this could be an indicator that participants would have preferred Rule A (which was not given as an option) or that they randomly chose between B and C.

Crucially, dance can also give rise to constituency and hierarchical grouping: when multiple properties that give rise to local grouping boundaries are present in a given sequence, higher-level grouping may be perceived by the onlooker, as illustrated by (50). Assuming that A is the grouping structure based on *both* change of speed and change of direction (Rule A in (49)), an onlooker may give more weight to one of these properties over the other, resulting in a higher-level group boundary as in A'; in this example, A'spsdown would result from weighting change of direction more than change of speed. The larger rectangles in (50) denote higher-level groups, and contrastively the smaller rectangles mark the local grouping structure. This is parallel to the notion of constituency present in language and music, which entails the usual properties of constituents, namely that they do not overlap with one another.

(50) Emergence of higher-level grouping in dance



Note that even in dance sequences that only contain one type of change, a higher-level grouping structure may emerge due to manipulation of the *strength* of this

change: for instance, if a dance sequence contains stronger and weaker changes of direction, all of them would give rise to a local grouping structure, out of which the strong ones (and only those) would trigger a higher-level grouping structure; see Charnavel (2016) for details.

To conclude this discussion, we can ask if we find *headedness* in dance, related to what we find in visual narratives, where narrative Peaks can ‘project’ to a higher level (Sect. 5.1.1), and in musical meter (illustrated in (39)) and time span reductions (illustrated in (40)), where a strong element projects to the next level (Sect. 5.2.1). A main difference between dance and the other artforms is that we can detect *categories* in visual narratives (e.g. Peak vs. Initial) and music (e.g. a tonic I chord vs. a dominant V chord, or a strong vs. weak beat). As of now, it is an open question whether similar categories can be posited for dance as well.

### 5.3.2 Dance semantics

The semantics of non-figurative / non-representational dance (e.g. some varieties of classical ballet, but also purely expressive dance in a club or a ballroom) is as yet an open question. But some narrative dances designed to tell a story might be easier to approach with semantic means. Patel-Grosz et al. (2018) probe for the presence of abstract meaning inferences in Bharatanatyam, a South Indian classical dance. Bharatanatyam is a type of narrative dance, thus closer to language and silent visual narratives than are other dance forms. In order to unearth how a dancer can encode meaning contrasts, Patel-Grosz et al. investigate coreference vs. disjoint reference in Bharatanatyam, through a series of motion capture production studies.

A sample stimulus is given below which consists of a context (= (51)) and target sentences (= (52)) (Patel-Grosz et al., 2022: 699–700). (52)a and (52)b are identical except for the number of characters involved: in (52)a, the artist sees the same man twice, hence a relation of coreference between the first and the second sentence; this is replaced in (52)b with a relation of disjoint reference.

(51) *Context*: An artist has designed a statue for a temple. She is at the temple, watching how people interact with the statue; the room is full of people.

(52) *Item 1*

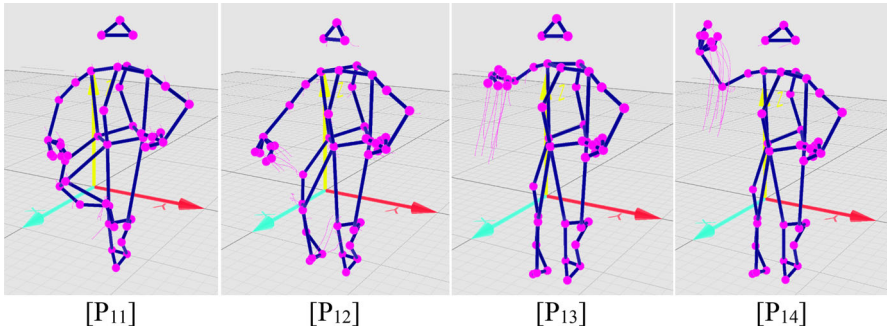
- a. The artist sees a strong man sitting on the ground.  
Then she sees that *the same man* is holding a spear. (coreference)
- b. The artist sees a strong man sitting on the ground.
- c. Then she sees that *another man* is holding a spear. (disjoint reference)

In a production experiment, the dancer performed a single, fluid movement in the coreferential condition, as shown in (53). In other words, coreference between  $P_{11}$  and  $P_{14}$  is not explicitly marked through specific mechanisms. Note that Bharatanatyam clearly operates with an iconic semantics; the dance posture that conveys that a man is sitting,  $P_{11}$ , or holding a spear,  $P_{14}$ , resembles that which it describes.

(53) Condition 1: coreference condition<sup>49</sup> (Patel-Grosz et al. 2022: 702)

The artist sees a strong man [P<sub>11</sub> sitting on the ground].

Then she sees that *the same man* [P<sub>14</sub> is holding a spear].

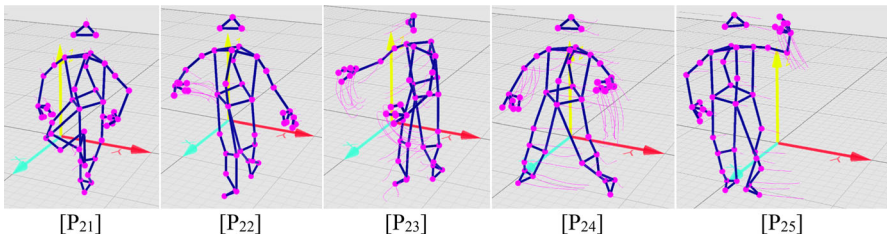


By contrast with the coreferential case, disjoint reference is marked, as illustrated in (54). As expected, the dancer begins in a seated position, identical to the coreference condition. The encoding of disjointness can be witnessed between P<sub>21</sub> and P<sub>25</sub>.

(54) Condition 2: disjoint reference condition (Patel-Grosz et al. 2022: 703)

The artist sees a strong man [P<sub>21</sub> sitting on the ground].

Then she sees that [P<sub>22+P23+P24</sub> *another man*] [P<sub>25</sub> is holding a spear]



The crucial observation, which Patel-Grosz et al. (2022) replicate in two follow-up studies, is that the dancer marks a new position in the visual space in P<sub>23</sub>, which she then assumes in P<sub>24</sub>, followed by the adoption of the spear-holding position in P<sub>25</sub>. The mechanism used by the dancer in P<sub>23</sub>-P<sub>24</sub>, marking the space followed by moving into it, is reminiscent of the use of reference tracking via loci, and action role shift (where a signer shifts their body, head position and/or eye gaze in order to assume the perspective of another referent) in sign language (cf. Padden, 1986; Lillo-Martin and Klima, 1990; Lillo-Martin, 1995; Quer, 2005; Sandler and Lillo-

<sup>49</sup> Videos with stick figures, from which these screen shots are taken were generated by the Qualisys Track Manager software (QTM, Qualisys AB, Gothenburg, Sweden). The complete set of videos from this study can be found at <https://www.youtube.com/playlist?list=PL0dCnZzwa9N5Q3qPHb9M79IITUb8A5Ww>.

Figure 1 corresponds to 0:10-0:12 in video “Set1 1A”: <https://youtu.be/jcvfNy24q5I?t=10>.

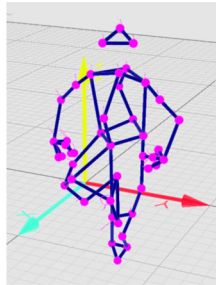
Figure 2 corresponds to 0:09-0:14 in video “Set1 1B”: <https://youtu.be/Xg3EmNGBDdQ?t=9>.

See Patel-Grosz et al. (2022) for the complete list of stimuli used in this study.

Martin, 2006; Herrmann and Steinbach 2009, 2012; see Davidson, 2015, for recent discussion of action role shift, and see Schlenker, 2017b for a recent survey article on loci). To avoid applying grammatical (linguistic) notions such as *loci* and *action role shift* to a phenomenon outside natural language, such as dance, Patel-Grosz et al. (2022) introduce the more neutral terms *indexical base*, defined as spatial positions that are assigned to a character in a narrative, and *action-performance*, defined as the demonstration of a character's actions from that character's viewpoint. The idea is that indexical bases and sign-language loci may owe to the same underlying cognitive processes, and that the same may be true for action-performances and sign-language role shift.

As discussed in Sects. 3.1 and 5.1.2, Greenberg (2011, 2013) and Abusch (2020) provide ways of defining truth in pictures and visual narratives, based on a generalized possible worlds model of information content (Abusch, 2020: 2), where the target forms rule out some possible situations and admit others. An illustration for a still from a dance sequence (e.g.  $P_{11}$  in (53),  $P_{21}$  in (54)) is given in (55). Truth in dance (as in other visual narratives) is defined in terms of whether a dance position  $P_n$  maps to a situation  $\sigma_n$  in the narrative; i.e. the dance position in (55) counts as satisfied by a fictional situation  $\sigma$  (i.e. "true" in  $\sigma$ ) if a sitting activity is taking place in  $\sigma$ .

- (55) Satisfaction conditions for dance position that describes a sitting activity (Patel-Grosz et al. 2022: 719)



a situation  $\sigma$  satisfies

only if in  $\sigma$  a person is sitting

A central question concerns the exact nature of the mechanism that we witness in  $P_{23}$  and  $P_{24}$ , since the movements of the dancer clearly create a visual discontinuity in the sense that the dancer's position and orientation change visibly. This visual discontinuity may contribute to the tracking of separate referents. There are thus at least two possible approaches (discontinuity-based referent management vs. indexical bases), and the respective hypotheses are stated in (56).



(56) **Hypothesis space for disjoint reference marking in dance** (slightly adapted from Patel-Grosz et al. 2022: 731)

To account for the change of position on part of the dancer when managing distinct variables ...

- a.  $H_1$  = ... visual discontinuity inferences (a type of visual iconicity) are sufficient, and no additional mechanisms, such as *indexical bases*, are needed.
- b.  $H_2$  = ... designated positions in space (*indexical bases*) are needed in addition to visual discontinuity marking.
- c.  $H_3$  = ... designated positions in space (*indexical bases*) are sufficient, and visual discontinuity marking is entirely uninformative with regards to coreference.

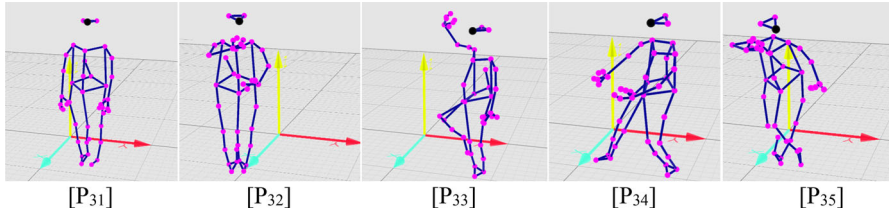
(56)a assumes that due to the absence of change of direction/orientation in the coreference condition, no grouping boundary arises. In the disjoint reference case, a grouping boundary is present due to the change of direction and orientation, which creates visual discontinuity. Such an approach is pursued by Patel-Grosz et al. (2018), but eventually rejected in Patel-Grosz et al. (2022). Unlike Charnavel (2019), who only assumes syntactic grouping (see Sect. 5.3.1), Patel-Grosz et al. (2018) propose that the syntactic groups are also interpreted as semantic groups, which correspond to events/subevents. This is parallel to the grouping structure assumed in music semantics, as alluded to in Sect. 5.2.2. Disjointness would then emerge by virtue of two referents contained in separate groups/events. It is worth noting that if (56)a is correct, the precise mechanism responsible for disjointness effects is underspecified. A plausible approach could be that visual iconicity gives rise to syntactic grouping as a byproduct of perceived discontinuity; this syntactic grouping then maps onto a semantic grouping into two (or more) events, which triggers an inference of disjoint reference (non-identity) between characters in these events. Patel-Grosz et al. (2022) reject such a view on the basis that it does not seem possible to predict the distinction between grouping boundaries that indicate disjoint reference vs. grouping boundaries that fail to do so; some additional mechanism is needed, in line with (56)b or (56)c.

The alternatives (56)b and (56)c assume that narrative dance uses a reference tracking mechanism that charts positions in space via *indexical bases*, reminiscent of the grammatical loci used in sign language (e.g., Liddell, 1990; Lillo-Martin and Klima, 1990). In order to test for the involvement of *indexical bases* in tracking disjoint referents, Patel-Grosz et al. (2022) develop the abovementioned production study and explore whether referents that occur earlier in the narrative can be picked up again at a later stage. This is answered in the affirmative, and *indexical bases* (positions in space) are operational for referent tracking in a similar fashion to the way loci are used in sign language.

An example sequence from a follow-up study is given in (57), adapted from Patel-Grosz et al. (2022: 713–714). This example introduces two referents (an eating child and a child holding a spear). The first referent (the eating child) is in a different position (base 1, in  $P_{32}$ ) from the second referent (base 2, in  $P_{33}$ ). When the first referent is picked up later in connection with a directional predicate (*watch*), the artist assumes base 1 again and faces towards base 2, as shown in  $P_{35}$ . This was reproduced across items and in a follow-up study with three different referents.

(57) **Retrieving referents by virtue of indexical bases**

[<sub>P31</sub> The artist]<sub>base0</sub> sees [<sub>P32</sub> a child eating a mango]<sub>base1</sub> outside the temple.  
 Then she sees [<sub>P33</sub> another child holding a spear]<sub>base2</sub>.  
 [<sub>(P34+)</sub>P35 *The eating child*<sub>base1</sub> watches *the child with the spear*<sub>base2</sub>].



What remains to be explored further is whether these positions in space exhibit other properties of sign language loci. An initial comparison in Patel-Grosz et al. (2022) shows that, at the very least, the arbitrariness of sign language loci is also found with indexical bases. Sign language loci do not obligatorily mark the actual (iconic) location of (fictitious or actual) physical objects or individuals in space, but can define arbitrary positions. Similarly, there is no implication that indexical bases in Bharatanatyam reflect the actual relative positions of referents in the narrative; Patel-Grosz et al. discuss a dance performance by Aishwarya Ravindran (AR) ([https://youtu.be/EJ1G\\_tvk59Q?t=605](https://youtu.be/EJ1G_tvk59Q?t=605)) where a base *a* is associated with the goddess Devi (at 10:05 of the video) while a base *b* is used to indicate the demon Mahisha (at 10:12). Patel-Grosz et al. observe that AR moves *through* the base of one referent while describing the actions of the other referent; this would be physically impossible if the bases marked their actual relative positions; it indicates that indexical bases, much like sign-language loci, function as abstract reference-tracking devices. It remains to be seen whether visual iconicity engages in grouping structure to create disjointness, in addition to the utilization of indexical bases, (56) b, or whether indexical bases are sufficient to track multiple referents, (56)c. Patel-Grosz et al. do not take a stance on whether  $H_2$  or  $H_3$  is correct, since indexical bases by their very nature entail visual discontinuity. The way indexical bases are utilized involves the marking of a position on stage, which the dancer moves into, in order to assume the perspective of a character in the narrative.

Regardless of whether (56)b or (56)c is correct, this exploration lends further support to something that we have seen for visual narratives in Sect. 5.1.2, and for music in Sect. 5.2.3. In all of these modalities, now also including dance, humans have the ability to track objects throughout a narrative, and this tracking can be modeled by virtue of variables. While grouping and constituency may be one property shared by these systems at the level of syntax, the tracking of denoted objects in a narrative may be one of their shared properties at the level of semantics.

#### 5.4 Interaction between visual and musical narratives

So far, we have discussed visual narratives (Sect. 5.1), music (Sect. 5.2) and dance (Sect. 5.3), where dance can also be treated as a type of visual narrative in some

cases, such as narrative dance, and minimally as a type of visual animation in other cases. We established that all of these media involve grouping structure both at a lower level and at a higher level, and that there are semantic commonalities, such as the use of variables. We can then ask how they interact, and, specifically, how does music interact with visual narratives? When studying music co-occurring with dance, films or cartoons, it is tempting to take one medium to just ‘repeat’ what the other says (e.g. with dance ‘interpreting’ the content of the music). But this is unlikely to be the right approach: both media have their own semantics, and the challenge is to understand how they can get aggregated.

The simplest hypothesis is that these contents are conjoined. If we focus on the highly simplified case of  $n$  pictures aligned with  $n$  musical events, the similarity between the semantic rules in (35) and (42) makes it possible to conjoin them (modulo some adjustments). We first need to take both kinds of sequences to be true of the same kinds of ( $n$ -tuples of) objects. So it will prove convenient to posit that pictures and musical events alike are true of eventualities, which may be states/situations or events. With this assumption, we can take a pictorial sequence  $\langle P_1, \dots, P_n \rangle$  aligned with a musical sequence  $\langle M_1, \dots, M_n \rangle$  to be true of eventualities  $\langle e_1, \dots, e_n \rangle$  just in case  $\langle P_1, \dots, P_n \rangle$  is pictorially true of  $\langle e_1, \dots, e_n \rangle$  and  $\langle M_1, \dots, M_n \rangle$  is musically true of  $\langle e_1, \dots, e_n \rangle$ . On this view, then, each medium makes its own contribution. There might be respects in which the two media ‘say the same thing’, but this is by no means a necessity.

In view of the foregoing discussion, each medium will come with variables, and thus anaphoric relations will serve to disambiguate cross-reference within and across media. An example (from Schlenker, 2022) might make this point concrete. (58) displays a simplified version of the beginning of Disney’s *Sorcerer’s Apprentice*, treated for simplicity as a sequence of 4 pictures combined with 4 musical events (in fact, these are 4 visual scenes combined with 4 musical themes). We first see the sorcerer with a musical theme A, then the apprentice with a musical theme B, then the sorcerer with the genie he just conjured, co-occurring with a modification of A ( $= A'$ ), then the genie alone, co-occurring with a modification of B ( $= B'$ ). In a simple analysis, each musical theme comes with a variable, hence the notations  $A[v_1]$ ,  $B[v_2]$ , for instance.

(58) **Four pictures from Disney’s *Sorcerer’s Apprentice* (below), with four musical events (above)** (Disney, Fantasia 1940; figure from Schlenker, 2022)

Video: <https://youtu.be/BR0Asbf2bxg>

Note that  $v_1$  denotes the sorcerer,  $v_2$  denotes the apprentice, and  $v_3$  denotes the genie.



The music alone might suggest that there is a total of two variables, one associated with A and A', and one associated with B and B' (just because of the similarity between A and A', B and B'). But when the music is aligned with the cartoon, things become more interesting. A and A' are naturally associated with the sorcerer, hence it is natural to posit that the variable  $v_1$  indeed appears in A and A', and also on the sorcerer in the first and third pictures. B is naturally understood to characterize the apprentice, hence a variable  $v_2$  appearing in B and in the second picture. However, B' is not associated with the apprentice, as one might expect on the basis of the music, but rather with the genie, and thus despite the similarity between B and B', it makes some sense to posit that the cross-reference between the image of the genie and B' is enforced by a third variable  $v_3$ . In this case, then, variables play a non-trivial role in establishing and tracking coreference both within and across media.

An example of dance-music interaction was alluded to in Sect. 5.2.3, example (48). In Michel Fokine's ballet choreography for Chopin's Mazurka op. 33 no. 2, we saw that in an [AB] [A'B'] sequence, different timbres in the music (orchestra for [AB] vs. oboe+flute for [A'B']) are mapped onto a choice between movements of the main ballerina [AB] vs. the other dancers [A'B']. While this was a case of complete correspondence between the music and the dance, we can easily imagine a situation where the other dancers only dance A', whereas B' is performed by a new dancer on the stage. This would then replicate the pattern that we just saw for the *Sorcerer's Apprentice*, in that the music would suggest two variables, whereas the dance sequence would introduce three. For the narrative dance sequences in Sect. 5.3, the question similarly arises whether an accompanying music would play a similarly supportive role in establishing and tracking variables.

While conjunction of visual and musical meaning might be a good starting point to analyze the interaction between visual and musical narratives, more sophisticated techniques from gesture semantics might prove useful as well. In a nutshell, it was proposed that, in some cases at least, music can enrich the meaning of visual sequences in the same way that gestures can enrich speech: by triggering cosuppositions (Schlenker, 2022).

The basic idea is illustrated in (59) by combining music with a simple gif representing Asterix drinking the magic potion, hitting a Roman soldier and leaving the room.

(59) *Context:* Asterix had an earlier encounter with a Roman soldier. Now he is faced with him once again.<sup>50</sup>

What will happen next? Will Asterix...

a.  WHISTLE



<https://youtu.be/aOyr8yTS6uY>

b.  WHISTLE



<https://youtu.be/BGElapUY4nw>

A light-hearted whistling tune accompanies either the entire scene (= (59)a) or just Asterix's departure (= (59)b). The entire excerpt is embedded in a question so as to tease apart at-issue content from presuppositions.

Unsurprisingly, the visual narrative provides information about the action and thus its main content is at-issue. Strikingly, the music adds inferences to the scene, but different ones depending on where the music appears. When the whistling tune co-occurs with the entire scene, (59)a, the whole sequence of actions is presented as light-hearted. When the whistling scene only co-occurs with the section in which Asterix departs, (59)b, the inference is different, along the lines of: *if Asterix leaves the room after drinking the potion and hitting the Roman soldier, his departure will be light-hearted*. Crucially, this inference exists despite the presence of a question, and it is conditionalized relative to the visual content it modifies. In these two respects, it behaves just like cosuppositions triggered by co-speech gestures. In the end, this behavior can be derived from a semantics for visual narratives along the lines of (35), combined with hypothesis that music semantics provides cosuppositional information about the part of the scene it co-occurs with (Schlenker, 2022).

A final point worth mentioning is that the timing of the music may also interact with the hierarchical structure of the cartoon. In terms of the narrative categories discussed in Sect. 5.1.1, the Asterix sequence in (59) can be mapped to Initial-Peak-Release, and the whistling either occurs during the Release, (59)b, or throughout the entire Arc, (59)a. In more complex narrative structures, e.g. with two Peaks, potentially, we might conceivably expect that the pairing of music with narrative categories is constrained in interesting ways, in that there may be Peak-specific types of music as opposed to, say, Release-specific types of music.

These examples raise two questions for future research. First, is it *always* the case that, when embedded in sentences, film and cartoon music makes the same kind of semantic contribution as co-speech gestures? Some caution is needed since the above discussion only cited a couple of examples. Second, does this behavior only arise when film and cartoon excerpts are embedded in sentences, or also in normal (unembedded) films and cartoons? The latter possibility is tantalizing but remote—

<sup>50</sup> Music: Billy Mowbray, Uke and Whistle. Gif: modified from <https://giphy.com/gifs/paf-asterix-ulCTAQ0E5ekV2>.

as was the case for the hypothesis that visual content on its own gets productively divided among some established slots of the inferential typology.

## 6 Semantics without phonology

The systems we have reviewed so far all have an overt realization: human language, primate calls, music, and visual narratives are instantiated in utterances we can perceive and parse, music we can hear, drawings we can see. We close by showing that the same primitives can be found in systems without an overt realization. This can be seen as a radical extension of the cognitive interpretation of semantics briefly discussed in Sect. 2.4: not only are semantic computations cognitively real; sometimes the human mind performs semantic computations on representations that have no overt instantiation at all.

Thought and reasoning allow us to manipulate representations, transforming models of what we know into models of conclusions we can draw from them, with crucial consequences for action. The tools of semantics can help illuminate this extraordinary human faculty. One starts with the description of a “language of thought” and a “logic of thought”. That is, we specify what the elementary building blocks (“words”) are and how they may be combined, and we can then derive new “sentences” from existing ones while preserving useful properties like truth. We outline concrete examples of insights that can be gained from this line of inquiry shortly. This semantics of thought doesn’t just radically extend the program of linguistic semantics; it might contribute to an understanding of its cognitive roots.

Still, there is an important methodological difference between the semantics of language and of thought. In the latter case, the absence of an overt realization forces us to reverse-engineer these systems through more indirect evidence: behavioral consequences must be used to infer what the atomic elements (lexical items) are. Developing language-free diagnostics makes it possible to export this enterprise to many systems, such as the reasoning and concept-formation faculties of non-human animals. In this sense, the semantics of thought and reasoning has been pursued for a long time. Ongoing research argues that these results can be regimented with the formal tools of semantics.

### 6.1 Content concepts and logical concepts

Linguistic semantics reveals constraints on minimal lexical meanings: similar concepts are found across languages, while some concepts are never or rarely found lexicalized. As we saw in Sect. 2.2.1, one such constraint is Gärdenfors’s connectedness, which has the result that no word may mean *mushroom or table*. Similarly, no word in no language is known to mean *none or all, few or many*, and a logical version of connectedness can derive this constraint as well, and it generalizes the old insight that primitive (i.e. lexical) determiners are either positive or negative.

Can similar constraints be displayed for language-free concepts? Consider the following thought experiment. You see sets of 5 objects each, and observe that groups with 2, 3 or 5 red objects behave in some way, while groups with 0 or 1 red

object do not. It is natural to infer that groups with 4 red objects will behave like the former: connectedness constrains generalizations drawn on the fly. Similarly, a connectedness-compatible generalization involving groups with at least 2 red objects will be easier to infer than a connectedness-incompatible generalization involving groups with 0, 1 or 4 red objects. Using tasks of this sort, we can arguably investigate people's language of thought without relying on their languages (Chemla et al., 2019a; Piantadosi et al., 2016), with interesting conclusions—including that the language of thought might make use of bound variables (e.g. Overlan et al., 2017). Still, how can we ensure that subjects are not relying on an inner linguistic monologue—which would make the final analysis about words? Researchers have typically made participants work with concepts that are not expressed in their language: even in such cases, concepts that obey connectedness are inferred more rapidly than ones that don't.

Importantly, such language-free tasks can be extended to non-linguistic participants, such as pre-linguistic infants and non-human animals, with striking results. To cite but one, baboons have been argued to obey a connectedness constraint just like human adults do: they infer connectedness-compatible generalizations faster than connectedness-incompatible ones (Chemla et al., 2019b). Thus a pervasive constraint in human content words and quantifiers arguably has roots in the conceptual behavior of rather distant non-human primates.<sup>51</sup>

## 6.2 Alternatives and attention

There are further interesting restrictions on cognitive modules without a phonology like thought, reasoning, or vision. We focus here on the role of (i) alternatives, which have played a key role in recent semantics/pragmatics but might also hold the key to some puzzles about reasoning, and (ii) attention as a facilitator of tractability for hard problems, operating by selecting among these alternatives. The findings have ramifications for the typology of inferences discussed above.

First of all, recent years have seen an increased interest in applying linguistic methodologies to reasoning problems in an attempt to illuminate the line between interpretation and reasoning. It is a truism both in linguistics and the psychology of reasoning that it is extremely dangerous to diagnose *failures* of reasoning without a solid grasp on the precise interpretation of the premises at hand. Indeed, the quest for absolving interpretations of well-known reasoning tasks has been of interest to psychologists since at least the 1980s.<sup>52</sup> Semanticists have begun contributing substantively to this indispensable line of research, for example using semantic

<sup>51</sup> In addition, conceptual constraints on call meanings might be useful in animal semantics (discussed in the Appendix). The heart of the matter is that any analysis of animal signals is greatly underspecified by available data. Constraints on animal concepts are plausibly constraints on signal meanings as well, and they might help adjudicate among competing theories.

<sup>52</sup> The conjunction fallacy of Tversky and Kahneman (1983) has been a particularly fertile ground for these investigations by psychologists (the fallacy arises when one assigns greater probability to  $p$  &  $q$  than to  $p$ ). For example, Dulany and Hilton (1991) apply Gricean pragmatics to articulate an alternative theory of the conjunction fallacy where ignorance inferences partly explain the phenomenon, and provide suggestive experimental evidence in favor of their account.

theories to explain framing effects (Geurts, 2013), or formal-pragmatics theories to explain rates of endorsement of apparent deductive fallacies (Picat, 2019).

Another key starting point is mental-model theory (Johnson-Laird, 1983), one of the most influential theories of human reasoning. Focusing on deductive reasoning, mental-model theory proposes a representational system that distinguishes between categorical premises that correspond to *exactly one mental model*, and another class of premises that instead propose a set of *alternative mental models* for consideration. Alternative mental models are generated by premises involving natural language disjunction (e.g. English *or*), and by reasoning-internal processes that allow humans to flesh out the full space of possibilities given a set of premises, at a potentially considerable computational cost.

For present purposes, the crucial idea behind alternatives in mental-model theory is that they represent not the possible states of affairs that we *know* to be compatible with the premises, but those that we happen to *be attending to*. Imagine you had the following two pieces of information as premises in a deductive problem: P1: *Either John speaks English and Mary speaks French, or Bill speaks German*; P2: *John speaks English*. Can you conclude with certainty that *Mary speaks French*? In a study of structurally identical examples, Walsh and Johnson-Laird (2004) show that over 85% of participants draw this conclusion. Yet it is a fallacy, for the following possible state of affairs makes the premises true and the conclusion false: John speaks English, Mary *does not* speak French, Bill speaks German. On mental model theory, the fallacious conclusion is a result of the alternative possibilities we are attending to. We *know* that the counterexample just presented is possible and invalidates the fallacious conclusion, but we are not *attending to* that possibility. The only possibilities we consider by default are the two possibilities expressed by the disjunctive premise, and those possibilities suggest that the only way to make the second premise *John speaks English* true is by making *Mary speaks French* true as well.<sup>53</sup>

<sup>53</sup> Two remarks are in order. Firstly, there is a scalar-implicature account of these particular illusory inferences with disjunction. As Mascarenhas (2014) shows, modern neo-Gricean and grammatical approaches to implicature predict that the first premise should be interpreted as “[John speaks English and Mary speaks French and Bill doesn’t speak German] or [Bill speaks German and John doesn’t speak English and Mary doesn’t speak French],” which provides an absolving interpretation for this apparent fallacy. Together with the second premise “John speaks English,” the conclusion “Mary speaks French” is in fact valid. Picat (2019) however shows that scalar implicature can only be part of the story, as psycholinguistic paradigms known to decrease the rate of scalar implicature calculation have only a very small effect on these fallacies, and have no effect on the broader paradigm of illusory inferences from alternatives. A multi-factor account holds the most promise: in these disjunctive inferences, two avenues conspire to create a very strong effect, one related to interpretive processes (scalar implicature), which analyzes the observed inference-making behavior as rational and in conformity with classical logic, the other a product of non-classical domain-general reasoning with alternatives (mental models or Koralus and Mascarenhas’s (2013) semantics-inspired variant).

Secondly, reasoning problems like these illusory inferences from disjunction have resisted accounts in psychology of reasoning frameworks other than mental models. For example, Sablé-Meyer and Mascarenhas (2022) prove that more recent probabilistic approaches to reasoning in the New Paradigm line (Oaksford and Chater, 2007) will only predict the observed fallacious conclusion to the extent that they predict an unobserved conclusion for the same deductive problem. Sablé-Meyer and Mascarenhas (2022) offer arguments in favor of unifying probabilistic approaches with the interpretive and model-theoretic commitments of work in the mental models tradition, but this direction of research is still in its earliest days.



Attention determines what alternative possibilities we consider when reasoning, and linguistic operators like disjunction drive attention. This is both a boon and a liability. On the one hand, it allows us to draw conclusions on the basis of a small space of alternative possibilities, reducing cognitive costs. On the other hand, it makes us vulnerable to fallacious reasoning.

What can semantics contribute to this research program? As mentioned, mental-model theory distinguishes between categorical premises and disjunctive premises that give rise to alternatives. Koralus and Mascarenhas (2013) show that this distinction is best regimented by incorporating semantic insights: semantic analyses of disjunction and several other operators have long made the same conceptual distinction as mental-model theory, and they thus offer off-the-shelf formal accounts of the representations of sentence meanings suitable to feed a mental-model-like reasoning module (e.g. the alternative semantics of Kratzer and Shimoyama (2002), based on Hamblin (1973); the inquisitive semantics of Groenendijk (2008), and Mascarenhas (2009); the truth maker semantics of Fine (2012)).

This discovery illustrates the fruitfulness of the interplay between semantics and reasoning. First, the formal landscape of theories of reasoning is enriched by this connection with linguistic semantics: semantic theories are characterized by their careful use of a broad palette of sophisticated formal systems, including non-standard logics developed in mathematics and computer science. The connection between semantics and reasoning allows us to import into reasoning brand new formal systems, offering new, useful, and insightful tools for theory building in the psychology of reasoning.

For example, Koralus and Mascarenhas's (2013) reimagining of mental-model theory offers a proof system that is associated with well-studied logical systems from the semantics literature. This proof system consists of a small number of operations that transform mental representations into other mental representations, much like, say, natural deduction systems provide rules for transforming formulas while preserving truth. Since Koralus and Mascarenhas (2013) were formalizing *intuitive* reasoning, their proposed operations on mental representations are not guaranteed to preserve truth in the general case. This proof-theoretic move allowed Koralus and Mascarenhas (2013) to prove meta-logical results with cognitive import. For example, they show that there is a derivation strategy (i.e. a reasoning strategy) in their system that guarantees logically valid conclusions, at the cost of exponential blowup of alternatives under consideration in the worst case. While the versions of mental-model theory due to Johnson-Laird and collaborators have typically come accompanied by open-source computer implementations, full-fledged soundness proofs of the kind just mentioned did not exist. The computer implementations of traditional mental-model theory did not constitute logics in the traditional sense, making soundness and completeness theorems hard to formulate and prove. Formal methodologies from semantics opened the door to proper meta-logical results in mental-model theory broadly conceived.

Second, the empirical scope of existing theories of reasoning increases in interesting ways. For example, the effects of attention and alternatives in the representations of disjunctive sentences discovered by mental-model theorists have been replicated in empirical domains where semantic theories have posited

disjunction-like alternatives: indefinites like *some* (Mascarenhas and Koralus, 2017) and epistemic operators like *might* (Bade et al., 2022; Johnson-Laird and Ragni, 2019; Mascarenhas and Picat, 2019). As the broader extent of the empirical phenomenon reveals itself, new and better constraints on our theories of reasoning emerge.

Finally, these results highlight the importance of cross-fertilization between theories of language and of thought, with the tantalizing possibility of broad unifications, as we have sketched here for the role of alternatives. Moreover, further cognitive modules can be brought into the fold in new ways. For instance, if effects of interpretation and reasoning with alternatives are at a deeper level about attention manipulation, then the need emerges to apply semantic methods to our study of visual attention as well.

It is interesting in this connection to return to the notions of implicatures, presuppositions, and supplements that we introduced before. One dimension of variation here concerns how attention is directed towards different parts of the world, or of a message in this case. We have argued that the division between these notions naturally arises beyond words (Sect. 4). In some cases at least, their pervasiveness and productivity in and beyond language might be due to how cognition is structured: the semantics of thought might yield the key to foundational issues in linguistic semantics.

## 7 Conclusion

This introduction outlined initial results of ‘Super Linguistics’, the application of tools and techniques from formal linguistics (and inspired by formal linguistics) to diverse non-standard representations, with the goal of contributing to a general formal theory of signs. A review of the achievements of formal linguistics in Sect. 2 (with a focus on syntax, semantics and pragmatics) was followed by two sections on iconic semantics (Sects. 3 and 4), with a particular focus on speech with gestures, and sign language semantics as an important point of reference. Section 5 focused on three human art forms (visual narratives, musical narratives, dance, and interactions between them), demonstrating the insights that we can gain from a syntactic and semantic analysis; most importantly, what all three art forms arguably share is hierarchical grouping in syntax and variables as a referent-tracking device in semantics. Finally, Sect. 6 showed how Super Linguistics can contribute to theories of language-free concepts and reasoning, by applying tools from linguistic semantics to the semantics of thought. Research in Super Linguistics has also demonstrated the applications of formal linguistic tools and techniques to animal communication, where syntax, semantics and pragmatics can each provide important analytical insights; this area is surveyed in the Appendix.

The benefits of applying linguistic methodology to non-standard objects are reciprocal: linguistics can contribute to objects of study beyond natural language by revealing structure and meaning principles in areas as diverse as gestures, music, dance and comics—as well as bird songs and monkey alarm calls (see the Appendix); at the same time, the study of these objects provides insights into

fundamental properties of cognition in humans and beyond, which contribute to our understanding of human language. Last, but not least, Super Linguistics offers an unprecedented expansion of the domains of application of linguistics as a field, with possible consequences for its broader scientific relevance.

The papers in this special issue cover, in particular, pictorial representations (Abusch and Rooth; Schlöder and Altshuler), music (Katz; Migotti and Guerrini), and linguistic approaches to body movement that include gestures (De Leon) and repetition-based physical exercises (Esipova). At the intersection of music and body movement, the special issue includes a contribution on dance (Charnavel), and at the intersection of gesture and pictorial representation, a contribution on emojis (Grosz, Greenberg, De Leon and Kaiser).

**Acknowledgments** We are very grateful to Neil Cohn, Gabe Greenberg, Patrick Georg Grosz, Naomi Francis, Jonah Katz and Emar Maier for extremely valuable written comments, and to Denis Bonnay for references. For helpful feedback and insightful discussion, we thank Vitor Nobrega, Cilene Rodrigues, audiences at the 2018 Crete Summer School of Linguistics workshop *Linguistic investigations beyond language: music, dance and visual narrative*; the X-Prag.de workshop *Gestures, body movement and primate linguistics*, ZAS Berlin; the *Musical meaning and beyond* workshop, NYU; the Introduction to Super Linguistics course at the 2020 New York–St. Petersburg Institute for Linguistics, Cognition and Culture; Abralin ao Vivo; the *Neubauer Collegium for Culture and Society*; Gabriel Greenberg’s graduate seminar on Non-Linguistic Representation; Crete Summer School of Linguistics course *Introduction to Super Linguistics*. Finally, we thank Emar Maier in his capacity as handling editor for *Linguistics and Philosophy*, for the countless hours he dedicated to the editing and reviewing process connected to this special issue. This paper is part of the Special Issue “Super Linguistics”, edited by Pritty Patel-Grosz, Emar Maier, and Philippe Schlenker.

**Funding** Open access funding provided by University of Oslo (incl Oslo University Hospital). This research was supported by funding from the Faculty of Humanities career development grant at the University of Oslo [PI: Patel-Grosz], and from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme [Grant Agreement No 788077, Orisem, PI: Schlenker]. Research was conducted in part at DEC, Ecole Normale Supérieure – PSL University. DEC is supported by grant FrontCog ANR-17-EURE-0017.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abner, N., Cooperrider, K., & Goldwin-Meadow, S. (2015). Gesture for linguists: A handy primer. *Language and Linguistics Compass*, 9(11), 437–449. <https://doi.org/10.1111/lnc3.12168>
- Abusán, M. (2011). Predicting the presuppositions of soft triggers. *Linguistics and Philosophy*, 34(6), 491–535. <https://doi.org/10.1007/s10988-012-9108-y>
- Abusch, D. (2013). Applying discourse semantics and pragmatics to co-reference in picture sequences. *Proceedings of Sinn und Bedeutung*, 17, 9–25.

- Abusch, D. (2020). Possible worlds semantics for pictures. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmermann (Eds.), *The Wiley Blackwell companion to semantics*. Wiley-Blackwell. <https://doi.org/10.1002/9781118788516.sem003>
- Abusch, D., & Rooth, M. (2017). The formal semantics of free perception in pictorial narratives. In A. Cremers, T. van Gessel, & F. Roelofsen (Eds.), *Proceedings of the 21st Amsterdam Colloquium* (pp. 85–94). ILLC, University of Amsterdam. <https://semanticsarchive.net/Archive/jZiM2FhZ/AC2017-Proceedings.pdf>
- Alikhani, M., & Stone, M. (2018a). Exploring coherence in visual explanations. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 272–277). IEEE. <https://doi.org/10.1109/MIPR.2018.00063>
- Alikhani, M., & Stone, M. (2018b). Arrows are the verbs of diagrams. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 3552–3563). Association for Computational Linguistics. <https://aclanthology.org/C18-1301>
- Alikhani, M., & Stone, M. (2019). “Caption” as a coherence relation: Evidence and implications. In *Proceedings of the Second Workshop on Shortcomings in Vision and Language* (pp. 58–67). Association for Computational Linguistics.
- Allwein, G., & Barwise, J. (Eds.). (1996). *Logical reasoning with diagrams*. Oxford University Press.
- Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.
- Aucouturier, J. J., Johansson, P., Hall, L., Segnini, R., Mercadié, L., & Watanabe, K. (2016). Covert digital manipulation of vocal emotion alter speakers’ emotional states in a congruent direction. *Proceedings of the National Academy of Sciences*, 113(4), 948–953. <https://doi.org/10.1073/pnas.1506552113>
- Bade, N., Picat, L., Chung, W., & Mascarenhas, S. (2022). Alternatives and attention in language and reasoning: A reply to Mascarenhas & Picat (2019). *Semantics and Pragmatics*. <https://doi.org/10.3765/sp.15.2>
- Barker, C., & Jacobson, P. I. (2007). *Direct compositionality*. Oxford University Press.
- Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2), 159–219.
- Bateman, J. A., & Schmidt, K. H. (2013). *Multimodal film analysis: How films mean*. Routledge.
- Bateman, J. A., & Wildfeuer, J. (2014). Defining units of analysis for the systematic analysis of comics: A discourse-based approach. *Studies in Comics*, 5(2), 373–403. [https://doi.org/10.1386/stic.5.2.373\\_1](https://doi.org/10.1386/stic.5.2.373_1)
- Beaver, D. I., Geurts, B., & Denlinger, K. (2021). Presupposition. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2021 edition). <https://plato.stanford.edu/archives/spr2021/entries/presupposition/>
- Beaver, D. I., Roberts, C., Simons, M., & Tonhauser, J. (2017). Questions under discussion: Where information structure meets projective content. *The Annual Review of Linguistics*, 3, 265–284. <https://doi.org/10.1146/annurev-linguistics-011516-033952>
- Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 51(3), 437–457.
- Charnavel, I. (2016). Steps towards a generative theory of dance cognition. Ms., Harvard University. <https://ling.auf.net/lingbuzz/003137>
- Charnavel, I. (2019). Steps towards a universal grammar of dance: Local grouping structure in basic human movement perception. *Frontiers in Psychology*, 10, 1364. <https://doi.org/10.3389/fpsyg.2019.01364>
- Chemla, E., Buccola, B., & Dautriche, I. (2019a). Connecting content and logical words. *Journal of Semantics*, 36(3), 531–547. <https://doi.org/10.1093/jos/ffz001>
- Chemla, E., Dautriche, I., Buccola, B., & Fagot, J. (2019b). Constraints on the lexicons of human languages have cognitive roots present in baboons (*Papio papio*). *Proceedings of the National Academy of Sciences*, 116(30), 14926–14930. <https://doi.org/10.1073/pnas.1907023116>
- Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th Boston University conference on language development* (pp. 157–168). Cascadilla Press.
- Chomsky, N. (1957). *Syntactic structures*. Mouton.
- Chomsky, N. (1980). *Rules and representations*. Columbia University Press.
- Cohn, N. (2012). *Structure, meaning, and constituency in visual narrative comprehension*. PhD thesis, Tufts University.

- Cohn, N. (2013). *The visual language of comics: Introduction to the structure and cognition of sequential images*. Bloomsbury.
- Cohn, N. (2015). How to analyze visual narratives: A tutorial in Visual Narrative Grammar. [http://www.visuallanguagelab.com/P/VNG\\_Tutorial.pdf](http://www.visuallanguagelab.com/P/VNG_Tutorial.pdf)
- Cohn, N. (2018). In defense of a “grammar” in the visual language of comics. *Journal of Pragmatics*, 127, 1–19. <https://doi.org/10.1016/j.pragma.2018.01.002>
- Cohn, N. (2020). Your brain on comics: A cognitive model of visual narrative comprehension. *Topics in Cognitive Science*, 12(1), 352–386. <https://doi.org/10.1111/tops.12421>
- Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology*, 65(1), 1–38. <https://doi.org/10.1016/j.cogpsych.2012.01.003>
- Cooper, R. (1983). *Quantification and syntactic theory*. Reidel.
- Coppola, M., Spaepen, E., & Goldin-Meadow, S. (2013). Communicating about quantity without a language model: Number devices in homesign grammar. *Cognitive Psychology*, 67(1–2), 1–25. <https://doi.org/10.1016/j.cogpsych.2013.05.003>
- Cumming, S., Greenberg, G., & Kelly, R. (2017). Conventions of viewpoint coherence in film. *Philosophers' Imprint*, 17(1), 1–29.
- Davidson, D. (1967). Truth and meaning. *Synthese*, 17(3), 304–323. <https://doi.org/10.1007/BF00485035>
- Davidson, K. (2015). Quotation, demonstration, and iconicity. *Linguistics and Philosophy*, 38(6), 477–520. <https://doi.org/10.1007/s10988-015-9180-1>
- Dingemanse, M. (2013). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, 6(10), 654–672. <https://doi.org/10.1002/lnc3.361>
- Dingemanse, M. (2015). Ideophones and reduplication: Depiction, description, and the interpretation of repeated talk in discourse. *Studies in Language*, 39(4), 946–970. <https://doi.org/10.1075/sl.39.4.05din>
- Dulany, D., & Hilton, D. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, 9(1), 85–110. <https://doi.org/10.1521/soco.1991.9.1.85>
- Ebert, C., & Ebert, C. (2014). Gestures, demonstratives, and the attributive/referential distinction. Handout of a talk given at *Semantics and Philosophy in Europe (SPE 7)*, Berlin, 28 June 2014.
- Esipova, M. (2019). *Composition and projection in speech and gesture*. PhD thesis, New York University.
- Ferrigno, S., Cheyette, S. J., Piantadosi, S. T., & Cantlon, J. F. (2020). Recursive sequence generation in monkeys, children, U.S. adults, and native Amazonians. *Science Advances*, 6, eaaz1002. <https://doi.org/10.1126/sciadv.aaz1002>
- Fine, K. (2012). A difficulty for the possible world analysis of counterfactuals. *Synthese*, 189(1), 29–57. <https://doi.org/10.1007/s11229-012-0094-y>
- Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303(5656), 377–380. <https://doi.org/10.1126/science.1089401>
- Frank, Annette (1996) *Context dependence in modal constructions*. Doctoral dissertation, University of Stuttgart.
- Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT Press.
- Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. MIT Press.
- Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, 440, 1204–1207. <https://doi.org/10.1038/nature04675>
- Geurts, B. (2013). Alternatives in framing and decision making. *Mind and Language*, 28, 1–19 <https://doi.org/10.1111/mila.12005>
- Goldin-Meadow, S. (2003). *The resilience of language*. Psychology Press.
- Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, 40, E46. <https://doi.org/10.1017/S0140525X15001247>
- Granroth-Wilding, M., & Steedman, M. (2014). A robust parser-interpreter for jazz chord sequences. *Journal of New Music Research*, 43(4), 355–374. <https://doi.org/10.1080/09298215.2014.910532>
- Greenberg, G. (2011). *The semiotic spectrum*. PhD thesis, Rutgers University.
- Greenberg, G. (2013). Beyond resemblance. *Philosophical Review*, 122(2), 215–287. <https://doi.org/10.1215/00318108-1963716>
- Greenberg, G. (2021a). Semantics of pictorial space. *Review of Philosophy and Psychology*, 12, 847–887. <https://doi.org/10.1007/s13164-020-00513-6>
- Greenberg, G. (2021b). The iconic-symbolic spectrum. Ms., UCLA. <https://ling.auf.net/lingbuzz/005787>
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66(3), 377–388. <https://doi.org/10.2307/2182440>

- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). Academic Press.
- Groenendijk, J. (2008). Inquisitive semantics: Two possibilities for disjunction. In P. Bosch, D. Gabelaia, & J. Lang (Eds.), *Proceedings of the Seventh International Tbilisi Symposium on Language, Logic and Computation* (pp. 80–94). Springer.
- Gutzmann, D., Matthewson, L., Meier, C., Rullmann, H., & Zimmermann, T. E. (Eds.) (2020). *The Wiley Blackwell companion to semantics*. Wiley-Blackwell.
- Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language*, 10, 41–53.
- Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. PhD thesis, University of Massachusetts at Amherst.
- Heim, I. (1983). On the projection problem for presuppositions. In M. Barlow, D. Flickinger, & M. Westcoat (Eds.), *Second Annual West Coast Conference on Formal Linguistics* (pp. 114–126). Stanford University.
- Heinz, J., & Idsardi, W. (2011). Sentence and word complexity. *Science*, 333(6040), 295–297. <https://doi.org/10.1126/science.1210358>
- Herrmann, A., & Steinbach, M. (2009). Visible context shifts: Quotation in sign languages. Talk presented at *42nd Annual Meeting of the Societas Linguistica Europaea (SLE 2009)*, University of Lisboa, Lisboa, Portugal.
- Herrmann, A., & Steinbach, M. (2012). Quotation in sign languages. In I. van Alphan & I. Buchstaller (Eds.), *Quotatives: Cross-linguistic and cross-disciplinary perspectives* (pp. 203–228). Benjamins.
- Hobbs, J. R. (1985). On the coherence and structure of discourse. Technical Report CSLI-85-37. Stanford University, Center for the Study of Language and Information.
- Hobbs, J. R. (1990). *Literature and cognition*. CSLI Publications.
- Horn, L. R. (1972). *On the semantic properties of logical operators in English*. PhD thesis, University of California Los Angeles.
- Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.
- Jackendoff, R. (2009). Parallels and non-parallels between language and music. *Music Perception*, 26, 195–204.
- Jacobson, P. (1999). Towards a variable-free semantics. *Linguistics and Philosophy*, 22(2), 117–184. <https://doi.org/10.1023/A:1005464228727>
- Jäger, G., & Rogers, J. (2012). Formal language theory: Refining the Chomsky hierarchy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598), 1956–1970. <https://doi.org/10.1098/rstb.2012.0077>
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge University Press.
- Johnson-Laird, P. N., & Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition*, 193, 103950. <https://doi.org/10.1016/j.cognition.2019.04.019>
- Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language* (pp. 277–322). Mathematical Centre.
- Kamp, H., & Reyle, U. (1993). *From discourse to logic*. Kluwer.
- Katz, J. (2022a). Metre, grouping, and event hierarchies in music: A tutorial for linguists. *Language and Linguistics Compass*, 16(9), e12472. <https://doi.org/10.1111/lnc3.12472>
- Katz, J. (2022b). Musical grouping as prosodic implementation. *Linguistics and Philosophy*. <https://doi.org/10.1007/s10988-022-09365-y>
- Keenan, E. L., & Westerståhl, D. (2011). Generalized quantifiers in linguistics and logic. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (pp. 859–910). Elsevier.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI Publications.
- Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z., Lameira, A. R., & Bohn, K. (2014). Animal vocal sequences: Not the Markov chains we thought they were. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20141370. <https://doi.org/10.1098/rspb.2014.1370>
- Klima, E. S., & Bellugi, U. (Eds.) (1979). *The signs of language*. Harvard University Press.
- Koralus, P., & Mascarenhas, S. (2013). The erotetic theory of reasoning: Bridges between formal semantics and the psychology of deductive inference. *Philosophical Perspectives*, 27, 312–365.
- Kratzer, A., & Shimoyama, J. (2002). Indeterminate pronouns: The view from Japanese. In Y. Otsu (Ed.), *Proceedings of the 3rd Tokyo Conference on Psycholinguistics* (pp. 1–25). Hituzi Syobo.

- Kuhn, J. (2016). ASL loci: Variables or features? *Journal of Semantics*, 33(3), 449–491. <https://doi.org/10.1093/jos/ffv005>
- Kuhn, J., & Aristodemo, V. (2017). Pluractionality, iconicity, and scope in French Sign Language. *Semantics and Pragmatics*. <https://doi.org/10.3765/sp.10.6>
- Lascarides, A., & Stone, M. (2009a). A formal semantic analysis of gesture. *Journal of Semantics*, 26(4), 393–449.
- Lascarides, A., & Stone, M. (2009b). Discourse coherence and gesture interpretation. *Gesture*, 9(2), 147–180.
- Lerdahl, F. (2001a). *Tonal pitch space*. Oxford University Press.
- Lerdahl, F. (2001b). The sounds of poetry viewed as music. *Annals of the New York Academy of Science*, 930(1), 337–354. <https://doi.org/10.1111/j.1749-6632.2001.tb05743.x>
- Lerdahl, F. (2013). Musical syntax and its relation to linguistic syntax. In M. A. Arbib (Ed.), *Language, music, and the brain* (pp. 257–272). MIT Press.
- Lerdahl, F. (2022). Notes towards a theory of musical narratives. Ms., Columbia University.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. MIT Press.
- Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.
- Liddell, S. K. (1990). Four functions of a locus: Re-examining the structure of space in ASL. In C. Lucas (Ed.), *Sign language research: Theoretical issues* (pp. 176–198). Gallaudet University Press.
- Liddell, S. K. (2003). *Grammar, gesture, and meaning in American Sign Language*. Cambridge University Press.
- Lillo-Martin, D. (1995). The point of view predicate in American Sign Language. In K. Emmorey & J. Reilly (Eds.), *Language, gesture, and space* (pp. 155–170). Lawrence Erlbaum Associates.
- Lillo-Martin, D., & Klima, E. S. (1990). Pointing out differences: ASL pronouns in syntactic theory. *Theoretical Issues in Sign Language Research*, 1, 191–210.
- Lillo-Martin, D., & Meier, R. P. (2011). On the linguistic status of ‘agreement’ in sign languages. *Theoretical Linguistics*, 37(3–4), 95–141. <https://doi.org/10.1515/thli.2011.009>
- Liuni, M., Ponsot, E., Bryant, G. A., & Aucouturier, J. J. (2020). Sound context modulates perceived vocal emotion. *Behavioural Processes*, 172, 104042. <https://doi.org/10.1016/j.beproc.2020.104042>
- Maienborn, C., von Stechow, K., & Portner, P. (Eds.). (2011). *Semantics: An international handbook of natural language meaning* (Vol. 1). De Gruyter Mouton.
- Maier, E. (2019). Picturing words: The semantics of speech balloons. In J. Schlöder, D. McHugh, & F. Roelofsen (Eds.), *Proceedings of the 22nd Amsterdam Colloquium* (pp. 584–592). ILLC, University of Amsterdam.
- Maier, E., & Bimpikou, S. (2019). Shifting perspectives in pictorial narratives. *Proceedings of Sinn und Bedeutung*, 23(2), 91–106. <https://doi.org/10.18148/sub/2019.v23i2.600>
- Mascarenhas, S. (2009). *Inquisitive semantics and logic*. MA thesis, ILLC, University of Amsterdam.
- Mascarenhas, S. (2014). *Formal semantics and the psychology of reasoning: Building new bridges and investigating interactions*. PhD thesis, New York University.
- Mascarenhas, S., & Koralus, P. (2017). Illusory inferences with quantifiers. *Thinking & Reasoning*, 23(1), 33–48. <https://doi.org/10.1080/13546783.2016.1167125>
- Mascarenhas, S., & Picat, L. (2019). ‘Might’ as a generator of alternatives: The view from reasoning. *Proceedings of Semantics and Linguistic Theory (SALT)*, 29, 549–561.
- McAuley, J. D., Wong, P. C., Mamidipaka, A., Phillips, N., & Margulis, E. H. (2021). Do you hear what I hear? Perceived narrative constitutes a semantic dimension for music. *Cognition*, 212, 104712. <https://doi.org/10.1016/j.cognition.2021.104712>
- Meyer, L. B. (1956). *Emotion and meaning in music*. University of Chicago Press.
- Migotti, L., & Guerrini, J. (2023). Linguistic inferences from pro-speech music. *Linguistics and Philosophy*. <https://doi.org/10.1007/s10988-022-09376-9>
- Migotti, L., & Zaradzki, L. (2019). Walk-denoting music: Refining music semantics. In J. Schlöder, D. McHugh, & F. Roelofsen (Eds.), *Proceedings of the 22nd Amsterdam Colloquium* (pp. 593–602). ILLC, University of Amsterdam.
- Montague, R. (1970a). English as a formal language. In B. Visentini (Ed.), *Linguaggi nella società e nella tecnica* (pp. 189–224). Edizioni di Comunità.
- Montague, R. (1970b). Universal grammar. *Theoria*, 36, 373–398.
- Montague, R. (1973). The proper treatment of quantification in ordinary English. In K. J. J. Hintikka, J. M. E. Moravcsik, & P. Suppes (Eds.), *Approaches to natural language* (pp. 221–242). Reidel.
- Morris, C. W. (1938). *Foundations of the theory of signs*. University of Chicago Press.

- Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.
- Overlan, M. C., Jacobs, R. A., & Piantadosi, S. T. (2017). Learning abstract visual concepts via probabilistic program induction in a Language of Thought. *Cognition*, *168*, 320–334. <https://doi.org/10.1016/j.cognition.2017.07.005>
- Padden, C. (1986). Verbs and role-shifting in American Sign Language. In C. Padden (Ed.), *Proceedings of the 4th National Symposium on Sign Language Research and Teaching* (pp. 44–57). National Association of the Deaf.
- Pasternak, R. (2019). The projection of co-speech sound effects. Ms., ZAS Berlin. <https://ling.auf.net/lingbuzz/004520>
- Pasternak, R., & Tieu, L. (2022). Co-linguistic content inferences: From gestures to sound effects and emoji. *Quarterly Journal of Experimental Psychology*, *75*(10), 1828–1843. <https://doi.org/10.1177/17470218221080645>
- Patel-Grosz, P., Grosz, P. G., Kelkar, T., & Jensenius, A. R. (2018). Coreference and disjoint reference in the semantics of narrative dance. *Proceedings of Sinn und Bedeutung*, *22*(2), 199–216.
- Patel-Grosz, P., Grosz, P. G., Kelkar, T., & Jensenius, A. R. (2022). Steps towards a semantics of dance. *Journal of Semantics*, *39*, 693–748. <https://doi.org/10.1093/jos/ffac009>
- Pesetsky, D. & Katz, J. (2009). The identity thesis for language and music. Ms., MIT. <https://ling.auf.net/lingbuzz/000959>
- Peters, S., & Westerståhl, D. (2006). *Quantifiers in language and logic*. Oxford University Press.
- Pfau, R., Salzmann, M., & Steinbach, M. (2018). The syntax of sign language agreement: Common ingredients, but unusual recipe. *Glossa: A Journal of General Linguistics*, *3*(1), 107. <https://doi.org/10.5334/gjgl.511>
- Pfau, R., Steinbach, M., & Woll, B. (Eds.) (2012). *Sign language: An international handbook*. De Gruyter Mouton.
- Piantadosi, S. T., Tenenbaum, J., & Goodman, N. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological Review*, *123*(4), 392–424.
- Picat, L. (2019). *Inferences with disjunction, interpretation or reasoning?* MA thesis, Ecole Normale Supérieure.
- Pietroski, P. M. (2018). *Conjoining meanings: Semantics without truth values*. Oxford University Press.
- Potts, C. (2005). *The logic of conventional implicatures*. Oxford University Press.
- Potts, C. (2015). Presupposition and implicature. In S. Lappin & C. Fox (Eds.), *The handbook of contemporary semantic theory* (2nd ed., pp. 168–202). Wiley-Blackwell.
- Pyllyshyn, Z. (2003). *Seeing and visualizing: It's not what you think*. MIT Press.
- Quer, J. (2005). Context shift and indexical variables in sign languages. *Proceedings of Semantics and Linguistic Theory (SALT)*, *15*, 152–168.
- Quer, J., Pfau, R., & Herrmann, A. (Eds.) (2021). *The Routledge handbook of theoretical and experimental sign language research*. Routledge.
- Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, *5*(1), 35–53. <https://doi.org/10.1080/17459737.2011.573676>
- Rohrmeier, M., Zuidema, W., Wiggins, G. A., & Scharff, C. (2015). Principles of structure building in music, language and animal song. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140097. <https://doi.org/10.1098/rstb.2014.0097>
- Sablé-Meyer, M., & Mascarenhas, S. (2022). Indirect illusory inferences from disjunction: A new bridge between deductive inference and representativeness. *Review of Philosophy and Psychology*, *13*, 567–592.
- Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge University Press.
- Schembri, A., Cormier, K., & Fenlon, J. (2018). Indicating verbs as typologically unique constructions: Reconsidering verb ‘agreement’ in sign languages. *Glossa: A Journal of General Linguistics*, *3*(1), 1–40. <https://doi.org/10.5334/gjgl.468>
- Schlenker, P. (2014). Iconic features. *Natural Language Semantics*, *22*(4), 299–356. <https://doi.org/10.1007/s11050-014-9106-4>
- Schlenker, P. (2017a). Outline of music semantics. *Music Perception: an Interdisciplinary Journal*, *35*(1), 3–37. <https://doi.org/10.1525/mp.2017.35.1.3>
- Schlenker, P. (2017b). Sign language and the foundations of anaphora. *Annual Review of Linguistics*, *3*, 149–177.



- Schlenker, P. (2018a). Gesture projection and cosuppositions. *Linguistics and Philosophy*, 41(3), 295–365. <https://doi.org/10.1007/s10988-017-9225-8>
- Schlenker, P. (2018b). Iconic pragmatics. *Natural Language & Linguistic Theory*, 36(3), 877–936. <https://doi.org/10.1007/s11049-017-9392-x>
- Schlenker, P. (2018c). Visible meaning: Sign language and the foundations of semantics. *Theoretical Linguistics*, 44(3–4), 123–208.
- Schlenker, P. (2018d). What is Super Semantics? *Philosophical Perspectives*, 32(1), 365–453. <https://doi.org/10.1111/phpe.12122>
- Schlenker, P. (2019a). Prolegomena to music semantics. *Review of Philosophy and Psychology*, 10(1), 35–111. <https://doi.org/10.1007/s13164-018-0384-5>
- Schlenker, P. (2019b). Gestural semantics: Replicating the typology of linguistic inferences with pro- and post-speech gestures. *Natural Language & Linguistic Theory*, 37(2), 735–784.
- Schlenker, P. (2020). Gestural grammar. *Natural Language & Linguistic Theory*, 38(3), 887–936. <https://doi.org/10.1007/s11049-019-09460-z>
- Schlenker, P. (2021a). The semantics and pragmatics of appositives. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmermann (Eds.), *The Wiley Blackwell companion to semantics*. Wiley-Blackwell. <https://doi.org/10.1002/9781118788516.sem110>
- Schlenker, P. (2021b). Iconic presuppositions. *Natural Language & Linguistic Theory*, 39(1), 215–289.
- Schlenker, P. (2022). Musical meaning within Super Semantics. *Linguistics and Philosophy*, 45(4), 795–872. <https://doi.org/10.1007/s10988-021-09329-8>
- Schlenker, P., & Chemla, E. (2018). Gestural agreement. *Natural Language & Linguistic Theory*, 36(2), 587–625. <https://doi.org/10.1007/s11049-017-9378-8>
- Schlenker, P., & Lamberton, J. (2019). Iconic plurality. *Linguistics and Philosophy*, 42(1), 45–108. <https://doi.org/10.1007/s10988-018-9236-0>
- Schlenker, P., & Lamberton, J. (2022). Meaningful blurs: The sources of repetition-based plurals in ASL. *Linguistics and Philosophy*, 45(2), 201–264. <https://doi.org/10.1007/s10988-020-09312-9>
- Schlenker, P., Lamberton, J., & Santoro, M. (2013). Iconic variables. *Linguistics and Philosophy*, 36(2), 91–149. <https://doi.org/10.1007/s10988-013-9129-1>
- Shimajima, A. (2015). *Semantic properties of diagrams and their cognitive potentials*. CSLI Publications.
- Shin, S. J. (1994). *The logical status of diagrams*. Cambridge University Press.
- Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua*, 117(6), 1034–1056. <https://doi.org/10.1016/j.lingua.2006.05.006>
- Stokoe, W. C. (1960). Sign language structure: An outline of the visual communication system of the American deaf. *Studies in Linguistics, Occasional Papers, Vol. 8*. University of Buffalo.
- Tarski, A. (1983 [1935]). The concept of truth in formalized languages. In J. Corcoran (Ed.), *Logic, semantics, metamathematics. Papers from 1923 to 1935 by Alfred Tarski*. Transl. by J. H. Woodger (pp. 152–278). 2nd edn., Hackett.
- Tarski, A. (1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*, 4(3), 341–376. <https://doi.org/10.2307/2102968>
- Thijssen, E. (1983). On some proposed universals of natural language. In A. ter Meulen (Ed.), *Studies in model-theoretic semantics* (pp. 19–36). Foris.
- Thomason, R. H. (Ed.) (1974). *Formal philosophy. Selected papers of Richard Montague*. Yale University Press.
- Tieu, L., Pasternak, R., Schlenker, P., & Chemla, E. (2017). Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks. *Glossa: A Journal of General Linguistics*, 2(1), 102. <https://doi.org/10.5334/gjgl.334>
- Tieu, L., Pasternak, R., Schlenker, P., & Chemla, E. (2018). Co-speech gesture projection: Evidence from inferential judgments. *Glossa: A Journal of General Linguistics*, 3(1), 109. <https://doi.org/10.5334/gjgl.580>
- Tieu, L., Schlenker, P., & Chemla, E. (2019). Linguistic inferences without words. *Proceedings of the National Academy of Sciences*, 116(20), 9796–9801. <https://doi.org/10.1073/pnas.1821018116>
- Tsoulas, G. (2017). Semantics in universal grammar. In I. Roberts (Ed.), *The Oxford handbook of universal grammar* (pp. 183–218). Oxford University Press.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293–315. <https://doi.org/10.1037/0033-295X.90.4.293>

- van Heijningen, C. A. A., de Visser, J., Zuidema, W., & ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences*, *106*(48), 20538–20543. <https://doi.org/10.1073/pnas.0908113106>
- Walsh, C., & Johnson-Laird, P. N. (2004). Coreference and reasoning. *Memory & Cognition*, *32*, 96–106.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung*, *4*, 301–350. [Engl. transl. ‘Laws of organization in perceptual forms’ published in W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). Routledge & Kegan Paul, 1938].
- Wildfeuer, J. (2014). *Film discourse interpretation: Towards a new paradigm for multimodal film analysis*. Routledge.
- Wiltschko, M. (2018). Discovering syntactic variation. In N. Hornstein, H. Lasnik, P. Patel-Grosz, & C. Yang (Eds.), *Syntactic structures after 60 years. The impact of the Chomskyan revolution in linguistics* (pp. 427–460). De Gruyter Mouton. <https://doi.org/10.1515/9781501506925-431>
- Zaradzki, L. (2021). *Les événements en sémantique linguistique et musicale*. Doctoral dissertation, Université Paris Diderot.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.