First and last as superlatives of before and after

Natural Language Semantics
https://doi.org/10.1007/s11050-024-09228-1

First and *last* as superlatives of *before* and *after*

Johanna Alstott[1]

Accepted: 30 December 2024

Abstract
First and *last* have been variously described as ordinals, superlatives, or both. These descriptions are generally not accompanied by extensive argumentation, and those who label *first* and *last* as superlatives do not present and argue for a particular decomposition. Thus, *first* and *last*'s status as ordinals vs. superlatives and their internal composition remain open issues. In this paper, I argue that *first* and *last* are superlatives, in particular the superlative forms of *before* and *after*. To argue that *first* and *last* are superlatives, I show that they pattern like superlatives and unlike ordinals (*second*, *third*, etc.) with respect to plurality, modifier choice, "modal superlatives" with *possible*, and the ordinal superlative construction. I next argue that the relations between *before* and *first* and between *after* and *last* show themselves overtly in many languages and in English paraphrases; furthermore, *first* and *last* semantically differ in ways that *before* and *after* have also been noted to differ. While I acknowledge one observation that *prima facie* counterexemplifies these claims, I argue that it constitutes a genuine counterexample only if one formalizes my decomposition of *first*/*last* using a standard Heimian (Heim in Notes on superlatives. Manuscript, MIT (1999)) entry for *-est*. The counterexample, which concerns the "upstairs *de dicto*" reading of superlatives, ceases to be an issue if one treats *before* and *after* as simplex and formalizes my decomposition using a Containment Hypothesis-inspired semantics (Bobaljik in Universals in comparative morphology: Suppletion, superlatives, and the structure of words, MIT Press, Cambridge, 2012) for *-est*.

Keywords Ordinals · Superlatives · Degree semantics · Temporal connectives · Before and after

1 Introduction

The small literature on the semantics of ordinal numbers (Bhatt 2006; Herdan and Sharvit 2006; Sharvit 2010; Yee 2010; Bylinina et al. 2014; Alstott 2023; Charnavel

✉ J. Alstott
jalstott@mit.edu

[1] Department of Linguistics and Philosophy, Massachusetts Institute of Technology, 32 Vassar St, Cambridge, 02139, MA, USA

Published online: 31 January 2025

2023) contains conflicting characterizations of *first* and *last*. Some treat them as regular ordinals on a par with *second*, *third*, etc. (Herdan and Sharvit 2006; Bylinina et al. 2014), describing *first* as a suppletive form for **one-th* and *last* as a non-decomposed item with an ordinal syntax-semantics. Others posit that *first* and *last* are not ordinal numbers at all but are instead superlative adjectives (Barbiers 2007; Charnavel 2023). Though there has been no thorough adjudication between the two positions for English, Barbiers (2007) argues for the latter perspective for Dutch *eerst* 'first'. Even so, the claim that *first* and *last* (or other languages' analogues) are superlatives has not been fleshed out: previous proponents of this claim have not extensively argued for a specific hypothesis about *first* and *last*'s internal structure. Barbiers (2007) and Charnavel (2023) note that *first* and *eerst* were once related to *fore* and the obsolete *eer* 'early', but because the decomposition of *first*/*eerst* is not their primary topic of interest, they do not show that these decompositions are active in the current language or give a formal analysis.

This work takes up the mantle of Barbiers (2007) and Charnavel (2023), giving several arguments that *first*/*last* are not ordinals and form a natural class with superlatives before arguing for a novel claim about what kind of superlatives *first*/*last* are: they are superlative counterparts of *before* and *after*.[1] In other words, *first* and *last* bear a similar semantic relation to *before* and *after* (respectively) as *tallest* bears to *taller*. The data leveraged to support this thesis are compatible both with analyses where *first*/*last* are morphological superlatives of *before*/*after* and with analyses where *first*/*last* are monomorphemic "lexical superlatives" whose entries are derived from those of *before*/*after*, but I opt to examine the prospects of a decompositional superlative analysis here. While one previous observation about "upstairs *de dicto*" readings (Bylinina et al. 2014) *prima facie* kills any decompositional superlative theory of *first*/*last*, I show that it becomes a non-issue just in case we treat *before* and *after* as simplex and formalize the decomposition using a Bobaljik (2012)-inspired semantics for *-est*.

I substantiate these claims in several steps, beginning by arguing that *first*/*last* are not ordinal numbers and instead form a natural class with superlatives. To make this argument, I discuss four ways in which superlatives differ from ordinals like *second*, *third*, etc., showing that *first*/*last* behave like superlatives and unlike ordinals in all cases. My first two arguments of this sort, inspired by Barbiers's (2007) arguments for Dutch, concern plurality and modifier choice. My other two arguments, which are novel, involve the ordinal superlative construction (Alstott 2023) and the "modal superlative" construction (Larson 2000; Schwarz 2005; Romero 2013).

Having argued that *first*/*last* form a natural class with superlatives to the exclusion of ordinals, I offer a variety of arguments that *first*/*last* bear a similar semantic relation to *before*/*after* as superlatives bear to comparatives. I start by showing that *first*/*last* are always paraphraseable as "before/after all others" and showing that many languages' terms for 'before' and 'first' or their terms for 'after' and 'last' have similar forms. I further demonstrate that *first* and *last* exhibit asymmetries in their truth conditions and veridicality that mirror well-known asymmetries between *before* and

[1] Though Charnavel (2023) mentions that *first* and *fore* are related, no work I know of posits a relation between *first* and *before* or between *last* and *after*.

after. I supplement these empirical arguments with a conceptual one: a theory that ties *first* to *before* and *last* to *after*—paired with Alstott's (2023) theory of *second*, *third*, etc.—helps explain an interpretive restriction on ordinals that must otherwise be stipulated.

After arguing that *first* and *last* are superlative counterparts of *before* and *after*, I examine the prospects of a decompositional implementation of this claim. I start by expounding the preferred way of formalizing the decomposition, which opts for the (standard) assumption that *before* and *after* are simplex. If *before* and *after* are simplex, the only way to say that *first*/*last* are decompositional superlatives of *before*/*after* is to use an entry for *-est* that can be the sister of a relational element like *before*/*after*. While standard Heim (1999)-style entries for *-est* do not fulfill this desideratum, Containment Hypothesis-inspired (Bobaljik 2012) entries for *-est* do: entries of this sort (e.g., Szabolcsi 2012; Coppock 2016) take a comparative relation *R* and add universal quantification (*R than all others*). If we assume this meaning for *-est*, we can say that *first* decomposes into *before* + *-est*, with *-est* taking the relational element *before* to produce the meaning *before all others*.

I argue that a theory where *first* = *before* + a Bobaljik-inspired *-est* and *last* = *after* + a Bobaljik-inspired *-est* captures a variety of data, including adverbial *first*/*last* (e.g., *Sal danced first/last*), adjectival *first*/*last* (e.g., *the first/last astronaut*), and *first*/*last*'s focus-sensitivity.

Having presented the preferred decompositional formalization of my thesis, I critique several alternative decompositions, most notably one that uses Heim's (1999) semantics for superlatives. On this analysis, *first* and *last* decompose into a gradable predicate + Heim's (1999) *-est*. Since *before* and *after* are not gradable predicates, the only way in Heim's system to encode a tight relation between *before* and *first* or between *after* and *last* is to say that *before*/*first* and *after*/*last* are comparative-superlative pairs for the same positive. For example, *before* and *first* might be *fore*+*-er* and *fore*+*-est*. I argue that this Heim-style theory has the fatal flaw that it—unlike the Bobaljik-style theory—is powerless to account for a difference between *first*/*last* and other superlatives regarding "upstairs *de dicto*" readings.

After comparing different decompositions (and briefly discussing a lexicalist alternative), I zoom out a bit, considering the question "why aren't *first* and *last* ordinals?" and giving a cross-linguistically informed answer inspired by Barbiers (2007).

The rest of this paper is structured as follows. Section 2 provides relevant background. Section 3 discusses past claims about *first*/*last*. Section 4 argues that *first*/*last* form a natural class with superlatives, while Sect. 5 argues that *first*/*last* are superlative counterparts of *before*/*after*. Section 6 expounds a Bobaljik-style decompositional formalization of this claim, while Sect. 7 considers alternatives. Section 8 discusses why *first*/*last* are not ordinals. Section 9 concludes.

2 Background

This section provides background on the two ongoing debates that play a role in this paper: (a) what is the decomposition of the superlative adjective? and (b) how similar are ordinals and superlatives?

2.1 Decomposing superlatives

The most basic analytical question in the study of superlatives is: what are the building blocks of a superlative? The literature contains two answers to this question.

The majority position in semantics (Szabolcsi 1986, Heim 1999, a.o.) dictates that superlatives decompose into a gradable predicate + *-est*. For example, *highest* decomposes into *high* + *-est*. On this theory, a comparative adjective (e.g., *higher*) is not one of the compositional pieces of a superlative.

The minority position in semantics (and majority position in morphosyntax), known as the Containment Hypothesis, following Bobaljik (2012), dictates that superlatives decompose into a gradable adjective + *-er* (the comparative morpheme) + *-est* (the superlative morpheme). For example, *highest* decomposes into *high* + *-er* + *-est*. The original motivations for this position stem from seminal work on the cross-linguistic morphosyntax of superlatives (Bobaljik 2012). While most semanticists have not incorporated Bobaljik's (2012) insights into their theories, some have (Szabolcsi 2012; Coppock 2016; Bumford and Sharvit 2022).[2]

Throughout most of this paper, I assume Bobaljik's (2012) position without further comment, formalizing my claims about *first*/*last* using a Containment Hypothesis-inspired entry for *-est* (Sect. 6). In Sect. 7.1, I consider whether one can formalize my claims using Heim's (1999) decomposition, arguing that this formalization faces a serious issue that the Bobaljik-style analysis does not. In Sect. 7.3, I touch on a third potential formalization, one where *first*/*last* are not decompositional superlatives but rather monomorphemic lexical items that have a superlative syntax-semantics.

2.2 How similar are ordinals and superlatives?

The semantic literature on ordinals starts from the observation that ordinals and superlatives are strikingly similar: they both are ambiguous between so-called "absolute" and "relative" readings (1a-b) (Szabolcsi 1986; Bhatt 2006), both license NPIs (2), and both license Bhatt's (2006) non-modal subject infinitives (e.g., the underlined clauses in (2a-b)).

(1) a. John climbed the highest mountain.
   i. **Absolute**: Out of all relevant mountains, John climbed the highest one.
   ii. **Relative**: John climbed a higher mountain than anyone else did.

   b. Ellie caught the first/(second) train.
   i. **Absolute**: Out of all relevant trains, Ellie caught the first/(second) one.
   ii. **Relative**: (With one exception,) Ellie caught a train before anyone else did.

[2] Stateva (2002) also claims that the LFs of superlatives contain a comparative-like element (labeled *-ER*). However, Stateva's (2002) theory does not entirely dovetail with the Containment Hypothesis, as *-ER* is not given the same meaning as the actual comparative morpheme *-er*.

(2) a. Yogi Berra was the best catcher to **ever** play in the MLB.
b. Donna Strickland is the third woman to **ever** win the Physics Nobel.

While these observations establish that ordinals and superlatives are related, previous theorists have disagreed about exactly how similar the lexical entries for ordinals and superlatives are. Most theories of ordinals (e.g., Bhatt 2006; Herdan and Sharvit 2006; Yee 2010; Bylinina et al. 2014) capture their similarities with superlatives by giving *n-th* and *-est* similar lexical entries. While these theories use entries for *n-th* that differ slightly from proposal to proposal, their entries all express virtually the same semantics: *n-th* is a function that, in *the second train* (3b), takes a comparison class C, takes *train*, and returns the singleton consisting of the train in C that is n-th in the relevant order. The entry in (3a), an amalgamation of existing proposals, formalizes this idea, where o(P) is a contextually salient order on the set P and a $>_{o(P)}$ b iff a outranks b in this order.

(3) a. $[\![\text{n-th}]\!] = \lambda C_{\langle e,t\rangle}.\ \lambda P_{\langle e,t\rangle}.\ \lambda x\text{: } x \in C \cap P \text{ and } |C \cap P| \geq n.$
$|\{y\text{: } y \in C \cap P \text{ and } y >_{o(C\cap P)} x\}| = n - 1$
b. $[\![\text{second train}]\!] = [\![[\text{two-th C}]\ \text{train}]\!] = \lambda x\text{: } x \in C \cap [\![\text{train}]\!] \text{ and } |C \cap [\![\text{train}]\!]| \geq 2.\ |\{y\text{: } y \in C \cap [\![\text{train}]\!] \text{ and } y >_{o(C\cap [\![\text{train}]\!])} x\}| = 1$

Both Heim (1999)-style and Bobaljik (2012)-style entries for *-est* have a semantic contribution very similar to that of (3a). These entries for *-est* take a comparison class argument (among others) and return the singleton containing the most highly ranked individual on a given dimension. Similarly, (3a) takes a comparison class argument (among others) and returns the singleton containing the n-th most highly ranked individual on a relevant dimension. Those who use (3a) exploit these parallels with *-est* to derive the similarities between ordinals and superlatives.

Contrary to proponents of (3a), Alstott (2023) claims that the parallels between ordinals and superlatives stem not from parallel lexical entries but from a covert superlative in ordinal DPs. To argue against entries like (3a), Alstott (2023) shows that the similarities between (3a) and *-est* leave proponents of (3a) unable to account for cases where a superlative and ordinal form a complex modifier, e.g., (4).

(4) Sarah climbed the eighth highest mountain.

Alstott's (2023) account of (4) involves an entry for *n-th* that takes *-est* as an argument; accordingly, the account posits a covert superlative in ordinal DPs where one is not present overtly. This covert superlative, Alstott (2023) argues, accounts for the similarities between ordinals and superlatives.

In Alstott (2023), I put no restrictions on what the covert superlative can be: I write that the covert superlative in *the second train* is *earliest* in a context where *the second train* means "the second train to arrive", while it is *leftmost* in a context where *the second train* means "the second train to the left in a line". In Sect. 5, I argue instead that *first* is the covert superlative in all ordinal DPs.

3 Previous claims about the status of *first* and *last*

Having reviewed relevant background on superlatives and ordinals, we can now discuss the issue relevant to both superlatives and ordinals that forms the subject of this

paper: namely, the internal structure of *first* and *last*. *First* and *last* have received varied labels in past literature. Some treat *first* and *last* as more similar to *second*, *third*, etc. than superlatives (Herdan and Sharvit 2006; Bylinina et al. 2014), decomposing *first* into **one-th* and treating *last* as a non-decomposed item with an ordinal syntax-semantics. Others suggest that *first* and *last*, unlike ordinals, are full-fledged superlative adjectives (Barbiers 2007; Charnavel 2023). Yet other work calls *first* "a superlative morpheme" (Sharvit 2010). Finally, Heycock (2005) calls *first/last* "both ordinals and superlatives".

The status of *first/last* is not the primary subject of most of these papers, however, so these characterizations of *first/last* are often not elaborated upon. The only previous claim about *first/last*'s internal structure that is central to an author's argument is Barbiers's (2007) claim that Dutch *eerst* 'first' is a superlative. Accordingly, only Barbiers 2007 addresses the status of *first* or *last* (or rather, *eerst*) in detail.[3] To argue that *eerst* is a superlative, Barbiers (2007) discusses three ways in which *eerst* patterns with superlatives and unlike other ordinals. Barbiers's (2007) first argument involves plurality: he claims that while superlatives and *eerst* can modify plural nouns, other ordinals cannot.

(5) de oud-ste/eerste/*vier-de boeken
the old-est/first/four-th books

Next, Barbiers notes that the *-ste* suffix in superlatives and *eerste* optionally reduces to *-st* in predicative position, while other ordinals that take *-ste* do not exhibit this optional shortening.

(6) Dit boek is het oud-st(e)/eerst(e)/achtst*(e)
this book is the old-est/first/eighth

Finally, Barbiers (2007) observes that superlatives and *first* take the modifier *aller-* 'very', while other ordinals do not.

(7) het aller-oud-ste / aller-eerste / *aller-honderd-ste boek
the very-old-est / very-first / very-hundred-th book

Although Barbiers (2007) supports his claim that *eerste* is a superlative, he does not give a formal semantic analysis, as his main topic lies elsewhere (see Sect. 8 for details). Barbiers (2007) briefly notes that *eerste*'s root is the obsolete *eer* 'early', but he does not focus on the question of whether *eer* is one of the compositional pieces of *eerste* in modern Dutch.[4] As such, many open questions remain regarding the structure of *eerste* and the structure of other languages' terms for 'first' (and 'last'), should they turn out to be superlatives too.

As an initial step towards addressing the question of the status of *first/last* cross-linguistically, I offer in this work an investigation of the status of English *first/last*,

[3]Charnavel (2023) elaborates briefly on her labeling of *first/last*, noting that the *st* in *first/last* is reminiscent of *-est*. Additionally, Sharvit (2015)—who does not commit to a position on *first*—briefly considers whether *first* contains *-est* because of certain similarities between *first* and superlatives (see Sect. 2.2 and fn. 8).

[4]Given the historical link between *eerste* and a root meaning *early*, one might wonder whether English *first* and *last* are suppletive superlatives of *early* and *late*, i.e., whether *first* = *earliest* and *last* = *latest*. I argue against this possibility in Sect. 7, fn. 36.

arguing for and formalizing a theory where they are superlative counterparts of *before* and *after*.

4 Ordinals vs. *first*/*last*/superlatives

To show that *first* and *last* are superlative counterparts of *before* and *after*, I start by arguing that they are not ordinals and instead form a natural class with superlatives. To do so, I show that variants of two of Barbiers' (2007) arguments about *eerste* extend to *first* and *last*; I supplement my Barbiers (2007)-inspired arguments with two novel arguments showcasing further ways in which *first*/*last* behave like superlatives and unlike ordinals.

My first Barbiers (2007)-inspired argument is similar to his plurality argument. Although both plural superlatives and plural ordinals are grammatical in English, plural superlatives are interpreted differently from plural ordinals. For a plural superlative sentence like (8a) to be true, A and B need not have arrived simultaneously.[5] But a plural ordinal sentence like (8b) is true and felicitous only if A and B arrived simultaneously (and after 10 other trains). Plurals with *first*/*last* act like plural superlatives rather than plural ordinals: (8c) does not entail that A and B arrived simultaneously.

(8) a. A and B were the earliest trains to arrive.
b. A and B were the eleventh trains to arrive.
c. A and B were the first/last trains to arrive.

Alstott (2023), the only previous author to analyze (8b), argues that its simultaneity inference stems from a presupposition of *-th* that *-est* lacks. If this analysis is on the right track, the fact that (8c) lacks a simultaneity inference suggests that *first*/*last* do not contain *-th* and lack a presupposition characteristic of ordinals.

My second Barbiers (2007)-inspired argument is similar to his modifier choice argument. Superlatives, *first*, and *last* can take the modifiers *very* and *absolute*, while *second*, *third*, etc. cannot (9a-c).

(9) a. The very/absolute best thing she told me was about you.
b. #The very/absolute second/third/fourth thing she told me was about you.
c. The very/absolute first/last thing she told me was about you.

Alstott (to appear), the first analysis of *very* that accounts for (9a) alongside cases like *very tall*, argues that *very* selects for morphemes that take a specially typed comparison class argument. If this analysis is on the right track, the contrast between (9b) and (9c) suggests that *first*/*last*'s comparison class argument is of the same type as that of superlatives and of a different type than that of *n-th*.

Let us turn to my two novel arguments, the first of which involves the "ordinal superlative" construction (Yee 2010; Alstott 2023). While ordinals can modify superlatives (10a), ordinals cannot modify other ordinals (10b).[6] However, ordinals can

[5] See Fitzgibbons et al. (2008) for more on cases like (8a).

[6] A reviewer notes that ordinals sometimes appear to modify ordinals: one can say *Caleb's second third strike* to mean "the third strike of the second at-bat where Caleb struck out". However, the syntax of *second*

modify *last* (*n-th (to) last*);[7] ordinals can modify *first* too in a context where *n-th-to-first* is the non-default way of interpreting *n-th* (10c). The fact that ordinals can modify superlatives and *first*/*last* but not other ordinals favors a theory where *first*/*last* are more akin to superlatives, especially on a theory where *second* in (10a-c) requires a superlative as an argument (Alstott 2023).

(10) a. Caleb was second-best.
b. #Caleb was second third.
c. A: I can't believe Emma was second (to) last.
B: Your list is upside-down. She was second-to-first!

My other novel argument concerns an ambiguity that superlatives with *possible* exhibit (Larson 2000; Schwarz 2005; Romero 2013): (11) can be used to talk about someone who is possibly a spy (11a), but it can also mean "Meg interviewed as smart a spy as possible for her to interview" (11b). Ordinals with *possible* possess only the former reading: while (12) can mean "Meg interviewed the fourth on a list of suspected spies", one cannot even construct a modal superlative paraphrase for (12) akin to (11b). Further support for the claim that (11) has a reading that (12) lacks comes from their felicity conditions. Because of its modal superlative reading, (11) is acceptable in a context where Meg only interviewed known spies. By contrast, (12) seems to require a context where Meg interviewed at least one "possible spy" (i.e., someone who is a suspected spy but not a known spy), suggesting the absence of a reading akin to (11b).

(11) Meg interviewed the smartest possible spy.
a. Modifier reading: Out of those who are possibly spies, Meg interviewed the smartest one.
b. Modal superlative reading: Meg interviewed as smart a spy as possible for her to interview.

(12) Meg interviewed the fourth possible spy.
a. Modifier reading: Out of those who are possibly spies, Meg interviewed the fourth one.
b. #Modal superlative reading

Unlike other ordinals, *first*/*last* possess a modal superlative reading (13).[8] To bring this reading out, suppose that Meg is tasked with interviewing a (known) spy. If Meg

third strike is [second [$_{NP}$ third strike]], whereas the syntax of *second-best train* and *second (to) last train* is [[$_{AP}$ second-best] train] and [[$_{AP}$ second (to) last] train]. Evidence for this constituency comes from the fact that *second-best*/*second last* can be predicative APs (10a, 10c), unlike *second third* (10b).

[7] *n-th last* is marked compared to *n-th-to-last*, but the Corpus of Contemporary American English (Davies 2008) contains many hits for *n-th last*, such as: *if raising taxes in a recession is the last thing you want to do, wouldn't raising taxes in a struggling economy teetering on a double-dip be the* ***second last*** *thing you'd want to do?* (source: LINK). Additionally, while it is more common to insert *to* in *n-th last* than in other *n-th* + *SUP* combinations, phrases like *second-to-lowest* do not register as ungrammatical to me and are attested: *the IT industry reported the second to lowest retention...* (source: LINK). So *last* and other superlatives do not categorically differ regarding *to*-insertion. I treat *to* as semantically vacuous.

[8] Sharvit (2015) notes the existence of modal superlative readings with *first*, but no prior work observes the contrast between *first*/*last* and other ordinals.

did the interview right away, we might say *Meg interviewed the first possible spy* to mean "she interviewed a spy as early as she could". But if Meg delayed finding a spy to interview until the last second, we might say *Meg interviewed the last possible spy*.

(13) Meg interviewed the first/last possible spy.
   a. Modifier reading: Out of those who are possibly spies, Meg interviewed the first/last one.
   b. Modal superlative reading: Meg interviewed a spy as early/late as possible.

Previous analyses posit a link between modal superlative readings and the presence of superlative morphology (Schwarz 2005; Romero 2013) or a link between modal superlative readings and the presence of an operator that allows *possible* to originate in a noncanonical position (Larson 2000). Regardless of which analysis one chooses, the data in (11)-(13) indicate that superlatives/*first*/*last* have a syntactic or semantic property in common that ordinals do not.

Having offered four arguments that *first*/*last* and superlatives form a natural class that excludes ordinals, I address a potential counterargument related to "upstairs *de dicto*" readings (Heim 1999). Heim (1999) notes that sentences like (14a) are acceptable in "upstairs *de dicto*" scenarios like (14). Bylinina et al. (2014) assert that ordinals (including *first*/*last*) are unacceptable in such scenarios, which is *prima facie* unexpected if *first*/*last* and superlatives form a natural class.[9]

(14) Context: John thought: "I want to take some train or other between 1pm and 2pm." Bill thought: "I want take some train or other between 3pm and 4pm." Mary thought: "I want to take some train or other between 5pm and 6pm." None of these people know anything about each other.
   a. John wanted to take the earliest train.
   b. ??/*John wanted to take the first train.
   c. ??/*Mary wanted to take the last train.
   d. ??/*Bill wanted to take the second train.

Although (14a-c) seem problematic for me, there is a way to account for (14a-c) while maintaining my claims about *first* and *last*. In particular, one could say that although *first*/*last* and *earliest* are all superlatives, they nonetheless differ in their internal structure in a way that explains the data in (14a-c). I develop a proposal along these lines in Sect. 7, where I exploit the fact that on my Bobaljik (2012)-style analysis, *first* (*before* + *-est*) and *last* (*after* + *-est*) differ from other superlatives (e.g., [[early -er] -est]) in that only the latter contain gradable adjectives.[10,11]

[9] Charnavel (2023) claims that the contrast diminishes upon adding a relative clause (*John/Mary/Bill wanted to take the first/second/last train that anyone wants to take*). On the basis of her judgments, Charnavel (2023) claims that the contrast in (14) relates to processing difficulty. I am not fully convinced of her judgments and thus do not adopt her explanation.

[10] One could also say that *first*/*last*, unlike other superlatives, are non-decompositional. See Sects. 5 and 7 for more on this possibility.

[11] Another potential counterargument concerns the judgment that while superlatives can be on either side of a numeral (*the two longest words*, *the longest two words*), *first*/*last* can only be to the left of a numeral

5 *Before* : *first* :: *after* : *last*

Having argued that *first*/*last* form a natural class with superlatives, I give six arguments that they should be thought of as superlative counterparts of *before* and *after*, i.e., that they bear a similar semantic relation to *before* and *after* as superlatives bear to comparatives. My first argument comes from paraphrases. Superlatives can always be paraphrased as universally quantified comparatives (Heim 1999; Bobaljik 2012). Thus, the equivalences in (15) hold. Both temporal (16) and nontemporal (17) uses of *first*/*last* can always be paraphrased with *before/after all others*, suggesting that *first* and *last* bear a semantic relation to *before* and *after* similar to the relation that *best* bears to *better*.[12]

(15) a. the best writer = the writer better than all others
b. the highest mountain = the mountain higher than all others

(16) a. The first battle = the battle before all others
b. The last day of school = the day of school after all others

(17) a. The first 2-digit number = the 2-digit number that is/comes before all others
b. Mel puts her mental health last ↔ Mel puts her mental health after all else

My next three arguments concern well-known semantic asymmetries between *before* and *after*; my claim is that *first* and *last* exhibit parallel asymmetries. The first

(*the first/last two words* vs. (?)*the two first/last words*). The COCA corpus (Davies 2008) has many hits for *two first/last NOUNs* (such as the following), so I suggest that both orders are actually grammatical. A reviewer provides other examples, e.g., *the four first words were "let there be light"*.

(i) a. In **the two first decades** of the century... (Source: Rudnitski 1995)
b. "...an edition of *Golden Hours* with **the two last pages** missing." (Source: Levin 2017)

[12] An anonymous reviewer provides the pairs in (i) and (ii) as potential cases where *first* is natural but *before all others* is less so.

(i) a. Mao is first among equals.
b. ??Mao is before all others among equals.

(ii) Context: Simone scored 10/10 points on the floor routine; everyone else scored fewer points.
a. Simone was first on the floor routine.
b. (?)Simone was before all others on the floor routine.

The reviewer concludes from (i)-(ii) that *first*, unlike *before*, can be used to talk about rank, but I think that these contrasts reflect extraneous factors. The contrast in (i) plausibly stems from the idiomatic status of *first among equals* (??*eighth among equals*); an idiom generally becomes less acceptable when you replace part of it with a synonym (*The compliment broke the ice* vs. #*The compliment broke the frozen water*).

I find that the contrast in (ii) satiates and is quite subtle. Additionally, (iib) improves if you replace *was* with *came*, so I do not think (iib) shows that *before* is inherently unable to be used in ranking contexts; rather, it shows that *first* is slightly more amenable to tracking the contextual ordering in copular sentences than *before*—perhaps because *first* is a superlative and thus has inherent access to a contextual variable that *before* does not. See Sect. 6.4 for relevant discussion.

asymmetry, originally observed by Heinamäki 1974, concerns the veridicality of *before*/*after*'s complement: *p before q* can be true even if *q* is never instantiated, while *p after q* entails that *q* occurred. For example, (18a) can be true even if Mozart never finished the Requiem, while (18b) entails that Mozart finished.[13]

(18) a. Mozart died before finishing the Requiem.
b. Mozart died after finishing the Requiem.

If *first*/*last* are superlative counterparts of *before*/*after*, we might expect there to be differences between *first*/*last* that mirror the veridicality asymmetry between *before*/*after*. And indeed there are. For example, the contrast between (19a) and (19b) follows directly from my hypothesis about *first*/*last* and the veridicality asymmetry between *before* and *after*. On my hypothesis, the first sentences in (19a) and (19b) are equivalent to *Sal sang before anyone else* and *Sal sang after everyone else*. So (19b) is contradictory and (19a) is not because *after everyone else sang* entails that everyone else sang, while *before anyone else sang* can be non-veridical.

(19) Sal, Meg, and Bill are taking turns singing.
a. Sal sang first. But no one else ended up singing because of a sudden tornado.
b. Sal sang last. #But no one else ended up singing because of a sudden tornado.

My next argument concerns a truth-conditional asymmetry between *before*- and *after*-sentences with atelics, originally observed by Anscombe (1964). To see this asymmetry, suppose that Ben and Fred sang from 6-9pm and consider (20a-b) relative to three scenarios for when Sal sang (Scenario A, 4-5pm; Scenario B, 7-8pm; Scenario C, 10-11pm). (20a) is true iff Sal sang before Ben started (Scenario A). By contrast, (20b) can be true either if Sal sang after Ben started (Scenario B) or if Sal sang after he finished (Scenario C).

(20) a. Sal sang before Ben sang.
b. Sal sang after Ben sang.

First and *last* show a parallel asymmetry. For example, consider (21a-b). If the relevant people are Sal, Ben, and Fred, (21a) is true iff Sal sang before Ben and Fred started singing (Scenario A). But (21b) can be true if Sal sang after Ben and Fred started singing (Scenario B) or if he sang after they finished (Scenario C).

(21) a. Sal sang first.
b. Sal sang last.

My fourth empirical argument concerns a truth-conditional asymmetry between *before*- and *after*-sentences with telics (originally observed by Heinamäki (1974)). To see this asymmetry, suppose that Meg and Will each climbed the mountain in three hours (2-5pm) and consider (22a-b) relative to three scenarios for when Josh climbed the mountain (Scenario A, 12-1pm; Scenario B, 3-4pm; Scenario C, 6-7pm). (22a)

[13] The examples in (18) come from Beaver and Condoravdi (2003).

is true if Josh climbed the mountain before Meg's climb started (Scenario A) and is also true if Josh climbed the mountain before Meg's climb finished (Scenario B). By contrast, (22b) is only true in Scenario C, where Josh climbed the mountain after Meg finished her climb.

(22) a. Josh climbed the mountain before Meg climbed the mountain.
b. Josh climbed the mountain after Meg climbed the mountain.

First and *last* show a parallel asymmetry, lending even more credence to the idea that *first* = *before all others* and *last* = *after all others*. For example, consider (23a-b). If the relevant people are Meg, Will, and Josh, (23a) is true if Josh climbed the mountain before the others started (Scenario A) and is true if he climbed the mountain before the others finished (Scenario B). But (23b) is only true in Scenario C.

(23) a. Josh climbed the mountain first.
b. Josh climbed the mountain last.

My final empirical argument relates to a cross-linguistic prediction my theory makes. If *first*/*last* are the superlative counterparts of *before*/*after* in English despite the lack of surface resemblance (as I have argued), then we might expect to find languages where the relationship between 'first'/'last' and 'before'/'after' is transparent. This prediction is borne out: many languages' terms for 'first' and 'before' or their terms for 'last' and 'after' have overt morphological and etymological links. Even in English, *first* and *before* share an historical link to *fore*: *first* was historically *fore*+-*est* (Oxford English Dictionary, s.v. "first (adj., adv., & n.2), Etymology", https://doi.org/10.1093/OED/9683472156), and *before* was the result of combining *fore* with *be-*, the prefix in *beset* and *besmear* (Oxford English Dictionary, s.v. "before (adv., prep., conj., n.), Etymology", https://doi.org/10.1093/OED/1018087550). In Italian, *first* and *before* are the same word (24).

(24) *Italian* (data from del Prete 2008):
a. Gianni arrivò **prima** che arrivasse Lea.
'Gianni arrived before Lea arrived.'
b. Leo è tornato il **prima** possibile
Leo is returned the before possible
'Leo returned at the first possible moment.'

Although *after* and *last* are not morphologically linked,[14] the terms for 'after' and 'last' in Hebrew share the same root. Similarly, the Mandarin term for 'last' is transparently built up from *zuì* 'most' and *hòu* 'after'.

(25) *Hebrew* (Omri Doron and Danny Fox, p.c.)
a. aχrej 'after', aχaron 'last'

[14] *Last* was an offshoot of *latest* that "[became] increasingly dissociated from *latest* in form and also in sense" (Oxford English Dictionary, s.v. "last (adv., adj., & n.4), Etymology", https://doi.org/10.1093/OED/3260300666). Given this etymology, one might wonder whether we should say that *last* = *latest* (instead of saying that *last* is the counterpart of *after*). I argue against this possibility in fn. 36.

(26) *Mandarin* (Yiyang Guo, p.c.)

a. **zuì-hòu** yī bān huǒchē
most-after one CL train
'the last train'

b. píngguǒ **zuì** hàochī
apple most delicious
'Apples are the most delicious.'

c. wǎnfān **hòu** lái zhǎo wǒ
dinner after come find me
'Come and find me after dinner.'

While these data validate the aforementioned cross-linguistic prediction of my account, note that I do not commit to the idea that *first*/*last* are derived from *before*/*after* in all languages. After all, some languages' terms for 'first' transparently derive from 'one', such as the Mandarin *dì-yī* (lit. ORD-one) and the Japanese *ichi-ban* (whose root means 'one'). It is *prima facie* implausible to treat these words as anything other than *one*+*-th*. Thus, I claim not that *first*/*last* are always derived from *before*/*after*, but rather that they are in English and many other languages. See Sect. 8 for further discussion.

The proposed analysis of *first*/*last* has not only empirical support but also a conceptual advantage: it offers a principled explanation of an interpretive restriction on *first*/*last* and ordinals that must otherwise be stipulated. To see what I have in mind, compare the range of possible meanings for ordinal superlatives with the range of possible meanings for *first*/*last* and "bare" ordinals. When an ordinal like *second* modifies an overt superlative, *second* can pick out the #2 element on a limitless variety of dimensions: the #2 element on the height dimension (27a), the #2 element on the "badly designed" dimension (27b), etc.

(27) a. The second-tallest building

b. The second most badly designed building

By contrast, when *second* does not modify an overt superlative, there is a strong preference to interpret it as picking out the #2 element on a temporal ordering or the #2 element on a proximate list; *first*/*last* are preferentially interpreted in the same ways. For example, there are two types of readings readily available for (28a-b): temporal readings (e.g., "the first/second/last building to be built")[15] and readings where (28a-b) refer to buildings on a list (e.g., a list of buildings in alphabetical order).[16] It is difficult to coerce (28a-b) into meaning something else; for example, one cannot coerce *the second building* into meaning *the second-tallest building* unless the context has a list of buildings ranked by height.

[15] See Heycock (2005), Bhatt (2006), Sharvit (2010), Bylinina et al. (2014), and Sects. 6.3-6.4 below for discussion on how infinitivals like *to be built* influence the ordering.

[16] I draw a distinction between temporal and list readings in this section for expository purposes, but I eventually claim that they are not formally distinct; see Sect. 6.4.

(28) a. The first/last building
b. The second building

Existing theories have to stipulate the interpretive restrictions on (28a-b). Bhatt (2006) and Bylinina et al. (2014) say that ordinals/*first*/*last* refer to a temporal "ranking function", but they do not explain why ordinals are inherently temporal. Sharvit (2010) and Alstott (2023) impose no restrictions on which orderings bare ordinals can be used to talk about, so they have nothing non-stipulative to say about the restrictions discussed above.

The proposed theory of *first*/*last*, paired with an extension of Alstott's (2023) theory of *second*, *third*, etc., paves the way for a more principled account. If *first*/*last* are superlative counterparts of *before*/*after*, we immediately account for why (28a) prefers temporal and "list" readings, as these are the readings that *before*/*after* have. *Before*/*after* can make claims about temporal order (29a) or the position of elements in a series (29b); hence, *first*/*last* can also be used to talk about time and lists. By contrast, *before*/*after* cannot compare two entities' heights, so *first*/*last* do not readily invoke a height-based ordering.

(29) a. Danielle kissed Alexis before/after Harry did.
b. 3 is before/after 4 on the number line.

If we adopt two further assumptions, we can extend this account of the interpretive restrictions on (28a) to (28b). The two assumptions are (i) Alstott's (2023) claim that *second*, *third*, etc. come with a covert superlative when one is not overt; and (ii) the novel claim that this covert superlative is always *first*.[17] On these assumptions, (28b) is equivalent to (30). *Second* picks out the #2 element on the ordering given by *first*, just like *second* in (27a) picks out the #2 element on the ordering given by *tallest*. Since *first* can only be used to talk about temporal or list orderings (for the reasons given above), *second* in (30) picks out the #2 element on a temporal or list ordering, as desired.

(30) The second ~~first~~ building

I conclude this section by making an important clarification. A reviewer observes that the arguments in Sects. 4 and 5 are compatible not only with an analysis where *first* and *last* are morphologically superlative (i.e., an analysis where they decompose into *X-est*) but also compatible with an analysis where *first* and *last* are "lexically superlative". On this sort of analysis, *first*/*last* are monomorphemic but have the distinctive syntactic-semantic properties of superlatives discussed in Sect. 4 and have meanings identical to the ones we would get if we decomposed *first*/*last* as morphological superlatives of *before*/*after*.

As the reviewer notes, we could argue for decomposition if we could establish that there is a subpart of *first*/*last* that scopes independently. But as discussed further in Sect. 7.2, scope considerations do not allow us to adjudicate between the lexicalist analysis and the version of the decompositional analysis laid forth in Sect. 6 below, leaving them on apparently equal footing.

[17] (ii) is not merely a stipulation; it explains why *n-th-to-first* feels redundant outside of special contexts like (10c).

For two reasons, I opt to focus on decompositional approaches in most of this paper. First, given that *first*/*last* bear the remnants of superlative morphology, I find it important to thoroughly consider whether an analysis where *first*/*last* contain *-est* is at all viable. Second, regardless of whether one prefers to encode the relationship between *first*/*last* and *before*/*after* in the syntax or the lexicon, decomposition is the easiest way to illustrate how to formally derive the meanings of *first*/*last* from those of *before*/*after*. However, I take the reviewer's point that a theory where *first*/*last* are lexically superlative—unlike a theory where they are ordinals and unlike certain implementations of the idea that they are morphologically superlative (see Sect. 7)—is a live alternative to the analysis I now provide.

6 A formal decompositional analysis

Having offered a variety of arguments for the position that *first* and *last* are superlative counterparts of *before* and *after* and opted to focus on a decompositional rather than a lexicalist version of this position, we now need to formalize the decomposition. We instantly face several important choice points, including (a) the choice between a Heim-style and a Bobaljik-style treatment of superlatives and (b) the choice of whether to treat *before* and *after* as internally simplex or as suppletive morphological comparatives. The proposal in this section starts from the (standard) assumption that *before* and *after* are simplex, an assumption that leads us to formalize the decomposition using a Bobaljik-style *-est*. In Sect. 7, I argue that Heim-style formalizations of the decomposition as well as Bobaljik-style formalizations that decompose *before* and *after* face serious issues that the preferred analysis does not.

The assumption that *before* and *after* are simplex leads us to formalize the decomposition using the Containment Hypothesis because if *before* and *after* are simplex, the only way to decompose *first* and *last* as superlatives of *before* and *after* is to use an entry for *-est* that can be the sister of a relational element like *before*/*after*. While Heim's *-est* does not meet this desideratum, Containment Hypothesis-inspired entries for *-est* do: such entries take a relation (comparative adjective, e.g.) and add universal quantification. In *tallest*, for instance, *-est* would attach to a silent comparative element and express universal quantification, with the result that *the tallest student* is the student who bears the *taller-than* relation to all others. Extending this analysis to the case at hand, the idea is that just as *-est* in *tallest* attaches to a comparative element and expresses universal quantification (*taller than all others*), *-est* in *first*/*last* attaches to *before*/*after* and expresses universal quantification (*before*/*after all others*).

6.1 Formal prerequisites

There are two prerequisites for an analysis where *first* = *before* + a Bobaljik-style *-est* and *last* = *after* + a Bobaljik-style *-est*: (i) a set of assumptions about *before*/*after* and (ii) a Bobaljik-style *-est* that works for superlatives other than *first*/*last* and allows us to avoid type mismatch in sentences containing *before*/*after* and *-est* together.

6.1.1 Assumptions about tense and *before*/*after*

There are two main theories of *before* and *after*: the quantificational theory of Anscombe (1964) and Krifka (2010) and the coercion-based theory of Beaver and Condoravdi (2003) and Condoravdi (2010). While the decomposition of *first*/*last* into *before-est*/*after-est* is compatible with a range of assumptions about *before* and *after*, I opt to utilize an Anscombe (1964)-style theory here.

Although my entries for *before*/*after* will be modeled on those of Anscombe (1964), I borrow from Condoravdi (2010) some background assumptions about tense and lexical aspect (because Anscombe 1964 does not make explicit assumptions). Following Condoravdi (2010), I make the simplifying assumption that there is one semantically interpreted tense operator in *before*- and *after*-sentences, an operator that existentially quantifies over times.[18] PAST, for example, takes a proposition p (a set of times) and returns true iff p has a member that precedes the evaluation time. Note that I introduce a type i for time-intervals and relativize the denotation function to a time-interval as well as an assignment function.

(31) $[\![\text{PAST}]\!]^{t,g} = \lambda p_{\langle i,t\rangle}.\ \exists t'\ [t' < t \text{ and } p(t') = 1]$

I assume that untensed clauses have truth values as their extensions and sets of times as their intensions. Like Condoravdi (2010), I represent untensed clauses in the metalanguage using the word *at*, e.g., (32a-b). $[\![\alpha]\!]^g_{\mathcal{C}}$ is α's intension, while $[\![\alpha]\!]^{t,g}$ is its extension (von Fintel and Heim 1997–2023).[19]

(32) a. $[\![\text{Ben sing}]\!]^g_{\mathcal{C}} = \lambda t'.$ Ben sings at t'

b. $[\![\text{Meg climb the mountain}]\!]^g_{\mathcal{C}} = \lambda t'.$ Meg climbs the mountain at t'

Condoravdi (2010) assumes that the meaning of an untensed clause depends on its *Aktionsart*. The intension of a stative like *Al be tired* contains all intervals throughout which Al is tired and is closed under the subinterval relation. The intension of an activity like *Ben sing* (32a) contains all the intervals throughout which Ben sings and is closed under the subinterval relation down to a small level of granularity. The intension of an accomplishment like *Meg climb the mountain* (32b) is a singleton containing the interval throughout which Meg climbs the mountain (it would be a non-singleton if Meg climbs the mountain multiple times). Finally, the intension of an achievement like *Meg leave* is the set of points at which Meg leaves.

Having introduced my assumptions about tense and untensed clauses, I now turn to my Anscombe (1964)-inspired account of *before* and *after*. To verify that we capture the two truth-conditional asymmetries from Sect. 5, I illustrate the theory with (20) and (22), repeated below as (33a-b).

(33) a. Sal sang before/after Ben sang.

b. Josh climbed the mountain before/after Meg climbed the mountain.

[18] Like Anscombe (1964) and Beaver and Condoravdi (2003) (as well as follow-up work), I ignore grammatical aspect. I expect that one can incorporate grammatical aspect into the theory without issue.

[19] $[\![\ ]\!]^g_{\mathcal{C}}$ is shorthand for $[\lambda t'.\ [\![\alpha]\!]^{t',g}]$.

I give (33a-b) the LFs in (34a-b), the derivations for which involve Anscombe's (1964) entries for *before/after* (35a-b).[20] $t'' \leq t$ means "t'' precedes or overlaps t".[21]

(34) a. [PAST [[Sal sing][before/after [Ben sing]]]]

b. [PAST [[Josh climb the mountain][before/after [Meg climb the mountain]]]]

(35) a. $[\![\text{before}]\!]^{t,g} = \lambda q_{\langle i,t\rangle}.\ \lambda p_{\langle i,t\rangle}.\ p(t) = 1$ and $\neg\exists t''\ [t'' \leq t$ and $q(t'') = 1]$

b. $[\![\text{after}]\!]^{t,g} = \lambda q_{\langle i,t\rangle}.\ \lambda p_{\langle i,t\rangle}.\ p(t) = 1$ and $\exists t''\ [t'' < t$ and $q(t'') = 1]$

I assume that *before/after* combine with their arguments via Intensional Functional Application (IFA) (von Fintel and Heim 1997–2023).[22] In other words, *before* and *after* combine with the intensions of their two clausal arguments. The sister of PAST (type ⟨it,t⟩) has an extension of type *t*, and hence PAST also combines with its sister via IFA. (36a) shows the predicted truth conditions for the version of (34a) with *before*, while (36b) shows the predictions for the version with *after*.

(36) a. $[\![[\text{PAST [[Sal sing][before [Ben sing]]]}]\!]^{t,g} = 1$ iff $\exists t' < t$: Sal sings at t' and $\neg\exists t''\ [t'' \leq t'$ and Ben sings at $t'']$

b. $[\![[\text{PAST [[Sal sing][after [Ben sing]]]}]\!]^{t,g} = 1$ iff $\exists t' < t$: Sal sings at t' and $\exists t''\ [t'' < t'$ and Ben sings at $t'']$

To verify that these truth conditions are correct, let us suppose again that Ben sang from 6-9pm and consider three scenarios for when Sal sang (Scenario A, 4-5pm; Scenario B, 7-8pm; Scenario C, 10-11pm). We correctly predict that *Sal sang before Ben sang* is only true in Scenario A, as that is the only scenario where we can find a Sal-singing time t' (4pm, let's say) such that there are no Ben-singing times before or overlapping t'. We also correctly predict that *Sal sang after Ben sang* is true in both Scenario B and Scenario C: in both scenarios, there is a Sal-singing time t' such that there is some Ben-singing time before t'.

Let us see what Anscombe's entries predict for our telic example. Remember that the matrix and embedded clauses denote singletons containing the runtime of Josh climbing the mountain and the runtime of Meg climbing it, respectively.

[20] A reviewer wonders how these entries can capture cases where *before/after*'s complement is type *i* (*before/after 2pm*). I assume that such cases involve applying Partee's (1986) IDENT type-shifter to $[\![2\text{pm}]\!]$. $\text{IDENT}([\![2\text{pm}]\!]) = \{2\text{pm}\}$; if we feed this singleton set to (35a-b), we get the correct truth conditions for *He left before/after 2pm*.

[21] Anscombe's (1964) entries are "non-uniform" in that they do not treat *before* and *after* as antonyms: (35a) and (35b) do not differ only in the temporal relation they encode ($<$ vs. $>$). If one places a high value on uniformity, one can easily recast my decomposition of *first/last* using Beaver and Condoravdi's (2003) uniform theory.

[22] I ignore possible worlds for simplicity, so I adopt the following temporal definition of IFA:

(i) **Intensional Functional Application**: If α is a branching node and $\{\beta,\gamma\}$ the set of its daughters, then, for any time *t* and assignment *g*: if $[\![\beta]\!]^{t,g}$ is a function whose domain contains $[\lambda t'.\ [\![\gamma]\!]^{t',g}]$, then $[\![\alpha]\!]^{t,g} = [\![\beta]\!]^{t,g}([\lambda t'.\ [\![\gamma]\!]^{t',g}])$.

(37) a. $[\![$[PAST [[Josh climb the mountain][before [Meg climb the mountain]]]]$]\!]^{t,g} = 1$ iff $\exists t' < t$: Josh climbs the mountain at t' and $\neg\exists t''$ [$t'' \leq t'$ and Meg climbs the mountain at t'']

b. $[\![$[PAST [[Josh climb the mountain][after [Meg climb the mountain]]]]$]\!]^{t,g} = 1$ iff $\exists t' < t$: Josh climbs the mountain at t' and $\exists t''$ [$t'' < t'$ and Meg climbs the mountain at t'']

While (37b) is correct, (37a) is not quite right. To see this, again suppose that Meg climbed the mountain from 2-5pm and consider three scenarios for when Josh's climb was (Scenario A, 12-1pm; Scenario B, 3-4pm; Scenario C, 6-7pm). Anscombe (1964) correctly predicts that *Josh climbed the mountain after Meg climbed the mountain* is only true in Scenario C, as that is the only scenario of the three where Meg's runtime precedes Josh's. However, Anscombe (1964) predicts that *Josh climbed the mountain before Meg climbed the mountain* is only true in Scenario A, whereas it is intuitively true in Scenarios A and B. To see this, note that unless Meg never climbed the mountain, the conditions in (37a) are only satisfied when Josh's runtime precedes Meg's runtime; there can be no overlap.

This issue for Anscombe (1964) is shared by Krifka's (2010) elaboration of Anscombe 1964. To remedy it, we can revise our entry for *before* as follows. $t''' \sqsubseteq t''$ means "t''' is a subinterval of t''."

(38) $[\![\text{before}]\!]^{t,g} = \lambda q_{\langle i,t\rangle}.\ \lambda p_{\langle i,t\rangle}.\ p(t) = 1$ and $\neg\exists t''$ [$\forall t''' \sqsubseteq t''$ [$t''' \leq t$] and $q(t'') = 1$]

Using (38), *Josh climbed the mountain before Meg climbed the mountain* comes out true iff there is a past time t' such that: (a) t' is the runtime of Josh climbing the mountain; (b) not all subintervals of Meg's runtime precede or overlap t'. Hence, the sentence is true in Scenario A and in Scenario B because in these scenarios, not all subintervals of Meg's runtime precede or overlap Josh's runtime. However, the sentence is false in Scenario C because every subinterval of Meg's runtime [2-5pm] precedes Josh's runtime [6-7pm]. I leave it to the reader to verify that adopting (38) does not undermine our predictions for other sentences (see also fns. 28 and 31).

This slightly tweaked version of Anscombe 1964 not only captures the truth conditions of *before*- and *after*-sentences but also their veridicality asymmetry. In other words, the theory correctly predicts that *p after q*, unlike *p before q*, entails *q*. The truth conditions for *p after q* have the form $\exists t'[p(t') = 1$ and $\exists t''$ [$q(''') = 1$ and ...]. These truth conditions can be satisfied only if *q* is instantiated at some time. By contrast, our truth conditions for *p before q* have the form $\exists t'$ [$p(t') = 1$ and $\neg\exists t''$ [$q(t'') = 1$ and ...]. These truth conditions are satisfied if *p* is instantiated and *q* is not.[23]

6.1.2 Assumptions about superlatives

Having laid out my assumptions about *before* and *after*, we now need some assumptions about superlatives other than *first/last*. As mentioned above, I adopt the Containment Hypothesis (Bobaljik 2012). Although there are semantic implementations

[23] Beaver and Condoravdi (2003) note that Anscombe (1964) needs to explain why "p before q and ¬q" is sometimes infelicitous (*#Ben met Sally before Fred did—in fact, Fred never met Sally at all*). Krifka (2010), a defender of Anscombe (1964), offers an explanation, arguing that *p before q* becomes more informative as the probability of *q* increases. See Beaver and Condoravdi (2003) for a different account.

of the Containment Hypothesis currently on the market (e.g., the phrasal analysis of Coppock 2016), I opt to use a novel Containment Hypothesis-inspired semantics instead. See Sect. 7.2 for discussion of Coppock's (2016) implementation of the Containment Hypothesis and an explanation of why I prefer my own.

On my implementation of the Containment Hypothesis, superlatives other than *first*/*last* have the "branching affix" structure (39a) rather than the "nesting" structure (39b) (terminology from Bobaljik 2012, p. 57). For example, *highest* decomposes into [high [-er -est]], *best* decomposes into [good [-er -est]], etc.

(39) a. [Adj [-er -est]]
b. [[Adj -er] -est]

Although Bobaljik (2012) uses (39b) in most of his discussion, he considers (39a) equally plausible for many languages, including English (p. 60). As discussed below, using (39a) allows us not just to compose *-er* and *-est* (in derivations with *highest*, *best*, etc.) but also to compose *before*/*after* and *-est* (in derivations with *first*/*last*).

I give gradable adjectives standard ⟨d,et⟩ entries (40a), give *-er* the standard two-place entry in (40b) (Heim 2000, a.o.), and *-est* the entry in (40c).[24]

(40) a. $[\![\text{high}]\!] = \lambda d.\lambda x.$ x's height $\geq$ d
b. $[\![\text{-er}]\!]^{t,g} = \lambda D'_{\langle d,t\rangle}.\ \lambda D_{\langle d,t\rangle}.\ \textsc{max}(D) > \textsc{max}(D')$
c. $[\![\text{-est}]\!]^{t,g} = \lambda R_{\langle dt,\langle dt,t\rangle\rangle}.\ \lambda C_{\langle dt,t\rangle}.\ \lambda D_{\langle d,t\rangle}.\ \forall Q\ [[Q \in C \text{ and } Q \neq D] \rightarrow R(Q)(D) = 1]$

Armed with these assumptions, I adopt a theory of superlatives like *highest*, *best*, *earliest*, etc. that is functionally identical to Heim's (1999) standard theory but assumes the Containment Hypothesis. On my theory, the sister of *high* in *highest* denotes the function in (41), i.e., the result of applying (40c) to (40b). Heim (1999) also assumes that the sister of *high* in *highest* denotes the function in (41): the only difference is that Heim (1999) treats (41) as the meaning of a simplex superlative operator (call it *-est*^{H}), whereas I treat it as the result of composing the comparative and superlative operators.[25]

(41) $[\![\text{-est}]\!]^{t,g}([\![\text{-er}]\!]^{t,g}) = \lambda C_{\langle dt,t\rangle}.\ \lambda D_{\langle d,t\rangle}.\ \forall Q \in C\ [Q \neq D \rightarrow \textsc{max}(D) > \textsc{max}(Q)]$

Since Heim (1999) and I both decompose a word like *highest* into (40a) + (41), I can adopt Heim's theory of sentences with these superlatives virtually wholesale. To see this, let us analyze the classic sentence (42). Recall from Sect. 2.2 that (42) is ambiguous between an "absolute" reading (which involves comparing mountains) and a "relative" reading (which involves comparing John to other mountain climbers).

(42) John climbed the highest mountain.
a. **Absolute**: Out of all relevant mountains, John climbed the highest one.
b. **Relative**: John climbed a higher mountain than anyone else did.

[24] Per Heim (2000), $\textsc{max}(D) = \iota d.\ D(d) = 1$ and $\forall d'\ [[D(d') = 1 \text{ and } d \neq d'] \rightarrow d > d']$.

[25] Heim (1999) also entertains a three-place entry for *-est*, but subsequent work has coalesced around the idea that (41) is superior (Romero 2013; Howard 2014; Charnavel 2023).

For Heim (1999), the LF for the relative reading involves focus on *John* and covert movement of [-est^H C] to the top of the clause (which leaves a degree trace and causes *the* to be interpreted as *a*).[26] I make the same assumptions but decompose *-est^H* into [-er -est] (as noted above). (43a) gives the proposed LF for the relative reading, and (43b) gives the predicted truth conditions.

(43) a. [[-er -est] C][∼C [λd. [PAST [[John]$_F$ climbs a d-high mountain]]]]

b. $[\![(43a)]\!]^{t,g} = 1$ iff
∀Q [[Q ∈ C and Q ≠ [λd. ∃t′ < t: John climbs a d-high mountain at t′]]
→
MAX([λd. ∃t′ < t: John climbs a d-high mountain at t′]) > MAX(Q)]

To see that these are the predicted truth conditions (and that they are correct), we need to say more about the comparison class and the treatment of focus. Like Heim (1999), I assume that C is a subset of the focus alternatives of [[-er -est] C]'s sister. To formalize this idea, I adopt Fox and Katzir's (2011) approach to focus, according to which the set of syntactic alternatives to an LF α, ALT-SYN(α), contains all LFs derivable from α by replacing focused constituents with constituents of the same type that are at most as complex, e.g.:

(44) ALT-SYN([λd. [PAST [[John]$_F$ climbs a d-high mountain]]]) =
{[λd. [PAST [John climbs a d-high mountain]]],
[λd. [PAST [Mary climbs a d-high mountain]]],
[λd. [PAST [Fred climbs a d-high mountain]]], . . . }

On the basis of ALT-SYN(), we can define the notion of an LF α's "semantic alternatives", ALT-SEM(α). (45) says that if α's intension is a temporal proposition (type ⟨i,t⟩), then ALT-SEM(α) is the set of intensions of members of ALT-SYN(α). Otherwise, ALT-SEM(α) is the set of extensions of members of ALT-SYN(α).

(45) For any LF α, time t, and assignment g:

$$\text{ALT-SEM}(\alpha) = \begin{cases} \{\lambda t'.[\![\beta]\!]^{t',g} | \beta \in \text{ALT-SYN}(\alpha)\} & \text{if } [\lambda t''.[\![\alpha]\!]^{t'',g}] \in D_{\langle i,t\rangle} \\ \{[\![\beta]\!]^{t,g} | \beta \in \text{ALT-SYN}(\alpha)\} & \text{otherwise} \end{cases}$$

(46) shows the semantic alternatives for the relevant node in (43a). Note that because the intension of this node is not ⟨i,t⟩, the semantic alternatives are the extensions of the syntactic alternatives rather than their intensions.

(46) ALT-SEM([λd. [PAST [[John]$_F$ climbs a d-high mountain]]]) =
{[λd. ∃t′ < t: John climbs a d-high mountain at t′],
[λd. ∃t′ < t: Mary climbs a d-high mountain at t′],
[λd. ∃t′ < t: Fred climbs a d-high mountain at t′], . . . }

At this point, one can constrain C in (43a) to be a subset of ALT-SEM([λd. [PAST [[John]$_F$ climbs a d-high mountain]]]). I execute this via an operator inspired by Rooth's (1992) ∼ operator, which I also call ∼.

[26] The idea that *the* becomes *a* in relative superlatives is standard, albeit *ad hoc* (Szabolcsi 1986; Heim 1999).

(47) $[\![\sim \mathrm{C}\ \alpha]\!]^{t,g}$ is defined iff $\mathrm{C} \subseteq$ ALT-SEM(α). When defined:

$$[\![\sim \mathrm{C}\alpha]\!]^{t,g} = \begin{cases} \lambda t'.[\![\alpha]\!]^{t',g} & \text{if } [\lambda t''.[\![\alpha]\!]^{t'',g}] \in \mathrm{D}_{\langle i,t\rangle} \\ [\![\alpha]\!]^{t,g} & \text{otherwise} \end{cases}$$

Due to the contribution of ∼C, the sister of [[-er -est] C] in (43a) presupposes that C is a subset of (46) and denotes $[\![$[λd. [PAST [[John]$_F$ climbs a d-high mountain]]]]$]\!]^{t,g}$ when defined.

Considering (43b) while keeping in mind that C is a subset of (46), we see that the sentence comes out true iff for any relevant $y \neq$ John, the height of the highest mountain John climbed exceeds the height of the highest mountain y climbed. These is the relative reading.[27]

Heim (1999) does not discuss how to derive absolute readings with (41), but Romero (2013) does. I assign the absolute reading of (42) her LF.

(48) [PAST [John climbed [the [λx. [[[-er -est] C] [∼C [λd. [x]$_F$ d-high mountain]]]]]]]

The object DP is composed in four steps. First, we merge [the [[high [[-er -est] C]] mountain]]. Then, [[-er -est] C] moves due to type mismatch with its sister *high* and leaves a trace, giving us [[-er -est] C][λd. [the [d-high mountain]]]. Third, we insert ∼C next to the surface sister of [[-er -est] C], giving us [[-er -est] C][∼C [λd. [the [d-high mountain]]]]. Finally, *the* moves and leaves a trace of type e, which receives focus.

As shown in (49a), C is a subset of {[λd. x is a d-high mountain] | x ∈ D_e}. Thus, the object DP denotes the unique mountain x such that the maximal height-degree possessed by x exceeds every other mountain's maximal height-degree. *John climbed the highest mountain* (on the absolute reading) asserts that John climbed this mountain, as desired.

(49) a. $\mathrm{C} \subseteq$ ALT-SEM([λd. [x]$_F$ [d-high mountain]]), i.e.:
$\mathrm{C} \subseteq$ {[λd. Everest is a d-high mountain],
[λd. K2 is a d-high mountain],
[λd. Kilimanjaro is a d-high mountain]...}

b. $[\![$the highest mountain$]\!]^{t,g} = \iota x.\ \forall Q\ [Q \in \mathrm{C}$ and $Q \neq$ [λd. x is a d-high mountain] → MAX([λd. x is a d-high mountain]) > MAX(Q)

[27] As noted by Howard (2014), there is one issue. Suppose John and Mary each climbed a 4,000ft mountain and Fred climbed a 3,000ft mountain. Here, *John climbed the highest mountain* is intuitively false on the relative reading. However, we predict it to be true: because [λd. John climbed a d-high mountain] = [λd. Mary climbed a d-high mountain] in this scenario, John and Mary will be associated with the same member of the alternative set, and thus we have no way of comparing John to Mary. To fix this issue, we can either introduce some modal machinery (as Howard 2014 does) or rewrite things so that C is an ordered tuple rather than a set: <[λd. John climbed a d-high mountain], [λd. Mary climbed a d-high mountain, ... >. If we opt for the latter route, we could add a conjunct to our entry for *-est* (40c) so as to ensure that its last argument, D, is not repeated in C, e.g., "C ≠ <D, ..., D, ... > and ∀Q..." I will not explicitly implement either option in the main text to simplify the derivations and exposition.

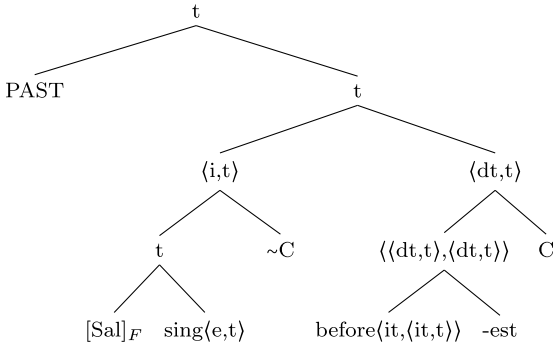
6.2 Temporal adverbial *first/last*

Having laid out assumptions about *before/after* and superlatives, I now present my analysis of *first* and *last*. This subsection and the next look at temporal adverbial (50a) and temporal adjectival (50b) uses of *first/last*. Section 6.4 turns to "list" readings.

(50) a. Sal sang first/last, Josh climbed the mountain first/last

b. The first/last astronaut

As discussed above, I treat *first* as *before-est* and *last* as *after-est*, meaning that I assign *Sal sang first* the LF below. Note that I make the additional assumption that type *i* (times) is a subtype of type *d*. Hence, *-est*'s first argument slot can be saturated by an element of type ⟨dt,⟨dt,t⟩⟩ (like *-er*) or an element of type ⟨it,⟨it,t⟩⟩ (like *before/after*). Similarly, *-est*'s second argument can either be ⟨dt,t⟩ or ⟨it,t⟩, etc. More generally, while an element of type *i* can saturate a function that is looking for something of type *d* (a more general type), the opposite is not true.



The definition of ALT-SEM given above says that because $[\![[\text{Sal}]_F \text{ sing}]\!]^{t,g}$ is of type *t*, ALT-SEM([[Sal]$_F$ sing]) is the set of temporal propositions in (51a). Using the definition of $\sim$ from above, we see that the sister of [[before -est] C] is defined iff C $\subseteq$ (51a); when defined, the sister of [[before -est] C] denotes the intension of [[Sal]$_F$ sing].

(51) a. ALT-SEM([[Sal]$_F$ sing]) = {[λt′. Sal sings at t′],
[λt′. Ben sings at t′],
[λt′. Fred sings at t′], … }

b. $[\![\sim$C [[Sal]$_F$ sing]$]\!]^{t,g}$ is defined iff C $\subseteq$ (51a).
When defined, $[\![\sim$C [[Sal]$_F$ sing]$]\!]^{t,g} = \lambda$t′. Sal sings at t′

-est takes *before* (type ⟨it,⟨it,t⟩⟩), C (type ⟨it,t⟩), and (51b) as arguments, and the intension of the result is fed to PAST. As shown in (52a), the sentence ultimately comes out true just in case *Sal sang before y sang* is true for every relevant individual $y \neq$ Sal. More precisely, the truth conditions state that at some past time t'', the temporal proposition [λt′. Sal sings at t′] bears the $[\![\text{before}]\!]^{t'',g}$-relation to every other temporal proposition in C. Given that C is constrained to be a subset of (51a), the truth

conditions boil down saying that at some past time t'', Sal sang at t'', and there is no time all of whose parts precede or overlap t'' at which anyone else sang.[28]

(52) a. $[\![\text{Sal sang first}]\!]^{t,g} =$
1 iff $\exists t'' < t$: $\forall Q$ [[$Q \in C$ and $Q \neq [\lambda t'.$ Sal sings at $t']$] $\rightarrow$
$[\![\text{before}]\!]^{t'',g}(Q)([\lambda t'.$ Sal sings at $t']) = 1$]

b. For any time t'' and temporal proposition Q,
$[\![\text{before}]\!]^{t'',g}(Q)([\lambda t'.$ Sal sings at $t']) = 1$ iff
Sal sings at t'' and $\neg\exists t'$ [$\forall t''' \sqsubseteq t'$ [$t''' \leq t''$] and $Q(t') = 1$]

To see that these truth conditions are correct, suppose again that Ben and Fred sang from 6-9pm and consider three scenarios for when Sal sang (Scenario A, 4-5pm, Scenario B, 7-8pm, Scenario C, 10-11pm). Assuming C has only the three members listed in (51a), we correctly predict that *Sal sang first* is only true in Scenario A, as that is the only scenario of the three where there is a past time t'' such that the relation in (52b) holds for every Q in the comparison class not equal to $[\lambda t'.$ Sal sings at $t']$. In other words, only in Scenario A is there a Sal-singing time t'' (4pm, let's say) such that there are no Ben-singing or Fred-singing times that precede or overlap t''.[29]

Using similar logic, we derive that *Sal sang after-est* (i.e., *Sal sang last*) is true in Scenarios B and C and false in Scenario A. This is because in Scenarios B and C, there is a past time t'' such that (53) holds for every Q in the comparison class not equal to $[\lambda t'.$ Sal sings at $t']$. In other words, only in Scenarios B and C is there a Sal-singing time t'' that comes after some of the Ben-singing and Fred-singing times.

(53) For any time t'' and temporal proposition Q,
$[\![\text{after}]\!]^{t'',g}(Q)([\lambda t'.$ Sal sings at $t']) = 1$ iff
Sal sings at t'' and $\exists t'[t' < t''$ and $Q(t') = 1]$

Turning to our telic example, we predict that *Josh climbed the mountain first* (54a) is true iff for every relevant y other than Josh, *Josh climbed the mountain before y climbed the mountain* is true. More precisely, it asserts that at some past time t'', the temporal proposition $[\lambda t'.$ Josh climbs the mountain at $t']$ bears the $[\![\text{before}]\!]^{t'',g}$-relation to every other member of C, where C has the structure in (54b).

(54) a. $[\![[\text{PAST } [[[\text{Josh}]_F \text{ climb the mountain}] \sim \text{C}][[\text{before -est}]\ \text{C}]]]\!]^{t,g} = 1$ iff
$\exists t'' < t$: $\forall Q$ [[$Q \in C$ and $Q \neq [\lambda t'.$ Josh climbs the mountain at $t']$] $\rightarrow$
$[\![\text{before}]\!]^{t'',g}(Q)([\lambda t'.$ Josh climbs the mountain at $t'])$ = 1]

b. $C \subseteq$ ALT-SEM($[[\text{Josh}]_F$ climb the mountain]), i.e.
$C \subseteq \{[\lambda t'.$ Josh climbs the mountain at $t']$,
$[\lambda t'.$ Meg climbs the mountain at $t']$,
$[\lambda t'.$ Will climbs the mountain at $t']$, ... $\}$

[28] Since *sing* is an activity, there are singing events that are virtually instantaneous. Thus, if there is no time all of whose parts precede or overlap t' at which anyone else sang, that means that no one other than Sal completed a minimal singing event until after t'.

[29] The predicted truth conditions for sentences with *first/last*, like the predicted truth conditions for other superlatives (see fn. 27), have an issue with "ties". (52a), for instance, is wrongly predicted to come out true if Sal and Ben sang from 4-5pm and Fred sang later; here, $[\lambda t'.$ Sal sings at $t'] = [\lambda t'.$ Ben sings at $t']$, and thus the truth conditions in (52a) will not compare Sal with Ben. We could fix this issue and the parallel issue with other superlatives in one breath by implementing either of the fixes from fn. 27.

Unpacking the $[\![\text{before}]\!]^{t'',g}$-relation as in (55), we see that we can state the truth conditions more precisely as follows: there is a past time t'' such that (a) t'' is the runtime of Josh climbing the mountain; and (b) for every relevant y other than Josh, there is no runtime of y climbing the mountain all of whose parts precede/overlap t''.

(55) For any time t'' and temporal proposition Q,
$[\![\text{before}]\!]^{t'',g}(\text{Q})([\lambda t'.$ Josh climbs the mountain at $t'])=1$ iff
Josh climbs the mountain at t'' and $\neg\exists t'$ $[\forall t''' \sqsubseteq t'$ $[t''' \leq t'']$ and $\text{Q}(t')=1]$

To see that these truth conditions are correct, suppose again that Meg and Will climbed the mountain from 2-5pm and consider three scenarios for when Josh's climb was (Scenario A, 12-1pm; Scenario B, 3-4pm; Scenario C, 6-7pm). Assuming the comparison class has only the three members listed in (54b), we correctly predict that *Josh climbed the mountain first* is only true in Scenarios A and B, as those are the scenarios where (55) holds for every Q in the comparison class not equal to $[\lambda t'.$ Josh climbs the mountain at $t']$. In other words, in both scenarios, not all subintervals of Meg and Will's runtimes precede or overlap Josh's runtime.

By similar logic, we predict (correctly) that *Josh climbed the mountain after-est* is only true in Scenario C, as that is the only scenario of the three where (56) holds for every Q in the comparison class not equal to $[\lambda t'.$ Josh climbs the mountain at $t']$. In other words, Scenario C is the only scenario where Josh's runtime follows Meg runtime and Will's runtime.

(56) For any time t'' and temporal proposition Q,
$[\![\text{after}]\!]^{t'',g}(\text{Q})([\lambda t'.$ Josh climbs the mountain at $t'])=1$ iff
Josh climbs the mountain at t'' and $\exists t'$ $[t' < t''$ and $\text{Q}(t')=1]$

Not only does the proposed theory capture the truth-conditional import of adverbial *first/last*, but it also captures *first/last*'s veridicality asymmetry. To see this, compare the predicted meanings of *Sal sang first* and *Sal sang last*. Due to the negation in $[\![\text{before}]\!]^{t,g}$, the predicted truth conditions of *Sal sang first* do not entail that anyone other than Sal sang. But the predicted truth conditions of *Sal sang last* entail that Sal and every other relevant y sang.

Finally, note that using focus machinery to derive the truth-conditional import of *first/last* is not *ad hoc*: *first/last* are known to be focus-sensitive (Bhatt 2006). For instance, Bhatt (2006) observes that sentences like *Danielle punched Alexis first/last* have different truth conditions depending on whether focus is placed on the subject or the object.

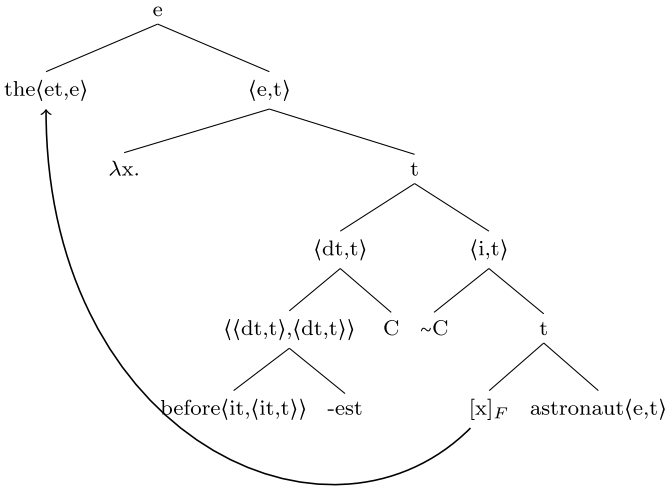
(57) a. $[\text{Danielle}]_F$ punched Alexis first/last.
$\rightsquigarrow$ Danielle was the first/last person to punch Alexis.
b. Danielle punched $[\text{Alexis}]_F$ first/last.
$\rightsquigarrow$ Alexis was the first/last person to be punched by Danielle.

In my system, the comparison class in (57a) is a subset of the set of temporal propositions $\{[\lambda t'.$ x punches Alexis at $t'] \mid x \in D_e\}$, and accordingly, (57a) makes a claim about the temporal order of Danielle punching Alexis vs. other people punching Alexis. By contrast, C in (57b) is a subset of the set of temporal propositions $\{[\lambda t'.$ Danielle punches x at $t'] \mid x \in D_e\}$, and accordingly, (57b) makes a claim about the temporal order of Danielle punching Alexis vs. Danielle punching other people.

6.3 Temporal adjectival *first*/*last*

My analysis covers not just temporal adverbial uses of *first*/*last* but also temporal adjectival uses. To see this, consider *the first/last astronaut*, whose intended temporal interpretation is paraphraseable as "the first/last person to be an astronaut". I assign *the first astronaut* the LF below and give *astronaut* the extension in (58).

(58) $[\![\text{astronaut}]\!]^{t,g} = \lambda x.$ x is an astronaut at t



To build this LF, we start by merging [[[before -est] C][~C [the astronaut]]]. Then, *the* moves and leaves a focused trace. Note that this LF is somewhat similar to (48), our LF for *the highest mountain*. The major difference is that while [-er -est] needed to move for type reasons in (48), *before-est* does not move in the LF above.[30] See Sect. 7 for more on this difference between *first*/*last* and other superlatives.

In the LF above, C is constrained to be a subset of ALT-SEM($[\![\text{x}]\!]_F$ astronaut]), i.e., a subset of the set of temporal propositions $\{[\lambda t'.$ x is an astronaut at $t'] \mid x \in D_e\}$.

(59) $[\![\text{C}]\!]^{t,g} \subseteq$ ALT-SEM($[\![\text{x}]\!]_F$ astronaut]), i.e.:
$[\![\text{C}]\!]^{t,g} \subseteq \{[\lambda t'.$ Sal is an astronaut at $t'],$
$[\lambda t'.$ Mary is an astronaut at $t'],$
$[\lambda t'.$ Jane is an astronaut at $t'], \ldots\}$

As a whole, $[\![\text{the first astronaut}]\!]^{t,g}$ denotes the unique x such that the temporal proposition $[\lambda t'.$ x is an astronaut at $t']$ bears the $[\![\text{before}]\!]^{t,g}$ relation to every other member of C. In other words, x is an astronaut at the evaluation time t, and there is no time preceding or overlapping t at which any other individual y was an astronaut.[31]

[30] Also, *highest* in (48) originates below the pre-movement position of *the*, whereas *before-est* originates above it. While motivated by types, this may explain why *first*/*last* tend to come higher in the DP than other superlatives (see fn. 11).

[31] Technically, the truth conditions say "there is no time **all of whose parts precede or overlap** t at which any other y was an astronaut." Since (58) is stative, though, "there is no *y-be-an-astronaut* time all of whose

(60) $[\![\text{the first astronaut}]\!]^{t,g} = \iota x.\ \forall Q\ [[Q \in C \text{ and } Q \neq [\lambda t'.\ x \text{ is an astronaut at } t']] \rightarrow [\![\text{before}]\!]^{t,g}(Q)([\lambda t'.\ x \text{ is an astronaut at } t']) = 1]$

Given our assumptions about PAST, we derive that a sentence like *Ben was the first astronaut* is true iff there is a past time t' such that Ben $= [\![\text{the first astronaut}]\!]^{t',g}$. In other words, Ben was an astronaut at t', and there is no time preceding or overlapping t' at which anyone else was an astronaut.

We currently account for some but not all temporal readings of adjectival *first/last*. To see this, consider *the first/last train*, which (as it stands) denotes the first/last entity to be a train. While this is one interpretation, this DP can also be interpreted in suitable contexts as *the first/last train to depart*, *the first/last train to pass us*, etc. To account for this ambiguity, I take a cue from Bhatt (2006) and Bylinina et al. (2014), positing that in *the first/last train*, *train* is the sister of a covert variable *P* (type ⟨e,t⟩) that combines with it via Predicate Modification and can optionally be expressed overtly by a non-modal subject infinitival (*to arrive*, *to pass by us*, etc.). When covert, $[\![P]\!]^{t,g}$ is contextually given.

On this theory, the shortened counterpart of *the first train to leave* receives LF (61a) and the meaning for *P* in (61b). In (61a), C is subject to the constraint in (61c), and as such, *the first train P* ends up denoting the train that left first, as desired. A variety of readings for adjectival *first/last* are derivable by similar means.

(61) a. the [λx. [[[before -est] C][∼C [[x]$_F$ [train P]]]]]

b. $[\![P]\!]^{t,g} = \lambda x.$ x leaves at t

c. $C \subseteq$ ALT-SEM([[x]$_F$ [train P]]), i.e.
$C \subseteq \{[\lambda t'.$ Train #1 is a train at t' and Train #1 leaves at $t']$,
$[\lambda t'.$ Train #2 is a train at t' and Train #2 leaves at $t']$,
$[\lambda t'.$ Train #3 is a train at t' and Train #3 leaves at $t']$, … $\}$

(62) $[\![\text{the first train P}]\!]^{t,g} =$
$\iota x.\ \forall Q\ [[Q \in C \text{ and } Q \neq [\lambda t'.\ x \text{ is a train at } t' \text{ and } x \text{ leaves at } t']] \rightarrow [\![\text{before}]\!]^{t,g}(Q)([\lambda t'.\ x \text{ is a train at } t' \text{ and } x \text{ leaves at } t']) = 1]$

Finally, my system can derive the fact that *first/last* exhibit the relative-absolute ambiguity ((1a), repeated below), albeit in a different way than with other superlatives.

(63) Ellie caught the first train.

a. **Absolute**: Out of all relevant trains, Ellie caught the first one.

b. **Relative**: Ellie caught a train before anyone else did.

Recall that I derive relative-absolute ambiguities for superlatives other than *first/last* via movement of [-er -est] (following Heim 1999). As discussed in Sect. 7, deriving *first/last*'s relative-absolute ambiguity in the same way is both empirically untenable and impossible in the current system. As such, I continue to assume a

parts precede or overlap t" is equivalent to "there is no point at which y is an astronaut that precedes or overlaps t."

movement theory of *highest*, *best*, etc. but adopt Bylinina et al.'s (2014) *in situ* approach to relative readings of *first*/*last*. On this theory, we derive absolute and relative readings for *the first train* by choosing different subsets of ALT-SEM($[[x]_F$ [train P]]) as our comparison class in (62). For example, suppose that there are three trains, #1-#3, and they leave in order of their number. Further suppose that Ellie took #2 and everyone else took #3. If C in (62) has the denotation in (64a), we get an absolute reading (which is false in this scenario): (62) denotes the first to leave out of all three trains (Train #1), and (63) is true iff Ellie took this train. If C is restricted to consider only the trains taken by a relevant person, i.e., #2 and #3 (64b), we get the relative reading: (62) denotes the first train to depart and be taken by a relevant person (Train #2), and (63) asserts that this train was taken by Ellie.

(64) a. C = {[λt′. Train #1 is a train at t′ and Train #1 leaves at t′],
[λt′. Train #2 is a train at t′ and Train #2 leaves at t′],
[λt′. Train #3 is a train at t′ and Train #3 leaves at t′]}

b. C = {[λt′. Train #2 is a train at t′ and Train #2 leaves at t′],
[λt′. Train #3 is a train at t′ and Train #3 leaves at t′]}

6.4 List readings

So far, I have focused on instances where *first*/*last* have a clearly temporal meaning rather than instances where *first*/*last* indicate position on a list, such as (65)-(66).

(65) The first/last letter in the alphabet

(66) Context: Ten trains are listed on a piece of paper. The one whose name is at the bottom is fast.

a. The last train is fast.

My theory straightforwardly extends to these "list" readings if we adopt Bylinina et al.'s (2014) perspective on these cases. Bylinina et al. (2014) observe that list readings can be thought of as having a temporal component: the existence of a list implies a process of "going over" the list (reading, recitation, etc.), a process which takes time. Because list readings have this temporal component, Bylinina et al. (2014) say, there is no categorical distinction between "temporal" and "list" readings.

Following Bylinina et al. (2014), we can formalize this idea using my earlier assumption that the head noun in *the last train* is the sister of a covert variable *P* that can optionally be expressed overtly by an infinitival clause (see Sect. 6.3). I claim that we get "temporal" readings of *the last train P* when $[\![P]\!]^{t,g} = [\![\text{to leave}]\!]^{t,g}$, $[\![P]\!]^{t,g} = [\![\text{to pass by us}]\!]^{t,g}$, etc., whereas the "list" reading exemplified in (66a) arises when $[\![P]\!]^{t,g} = [\![\text{to appear if I scan the list from top to bottom}]\!]^{t,g}$. With this *P*, the subject DP in (66a) is equivalent to *the last train to appear if I scan the list from top to bottom*, the semantics for which involves temporal comparison between trains in terms of when they would appear to the speaker if the speaker visually went over the list. Similarly, we can think of (65)—on the relevant reading—as being equivalent to *the first letter in the alphabet to be read if I recite the alphabet in canonical order*.

Curiously, adverbial *first/last* do not show the ambiguity of their adjectival counterparts. For example, *[Sal]$_F$ sang first* only has the meaning "Sal sang before everyone else sang", *Danielle punched [Alexis]$_F$ last* only means "Danielle punched Alexis after she punched everyone else", etc.[32] Thus, contextually salient lists do not factor into the truth conditions of sentences with adverbial *first/last* outside of cases like *"a" comes first in the alphabet*, where the verb itself (presumably) lexicalizes the meaning "is encountered at t if one goes over the relevant list in the relevant way". This contrast between adjectival and adverbial *first/last* follows if we assume that domain restriction variables like P occur in the nominal domain and not the verbal domain.

The above account of *first/last*'s list readings extends to list readings of *before/after*. List readings of *before/after* are most easily accessible in sentences with *come* (67a), but they are also accessible in copular sentences (67b).

(67) a. "a" comes before "b".

b. "a" is before "b".

I assume that there is an elided predicate in the *before*-clause in such sentences (Overfelt 2021), e.g., (67b) = *"a" is before "b"* ~~*is*~~. If we make this assumption, (67a) can get a treatment parallel to the above treatment of *"a" comes first*. For (67b), note that there is a type issue if we assume a null copula: *before* needs two propositional arguments, but neither clause has a non-vacuous predicate. I assume that although variables like P generally are restricted to the nominal domain, P is emergency-inserted in (67b) to resolve the mismatch (*"a" is P before "b"* ~~*is*~~ P). On the relevant reading of (67b), $[\![\mathrm{P}]\!]^{t,g} = [\![\mathrm{comes}]\!]^{t,g}$.[33,34]

7 Comparison with alternatives

Having presented one way of formalizing my claims about *first* and *last*, I now consider alternative formalizations. Section 7.1 critiques alternative decompositions of *first/last*, i.e., superlative decompositions that do not treat *first* as *before+-est* and *last* as *after+-est*. Section 7.2 considers a theory where we still treat *first* as *before-est* and *last* as *after-est* but adopt Coppock's (2016) phrasal analysis of *-est*. Finally, Sect. 7.3 considers a theory where *first/last* are "lexical superlatives".

[32] *John ranked [Mary]$_F$ first* seems to be a counterexample: it has a temporal reading ("John decided on [Mary]$_F$'s position on the list first") and a list reading ("John put Mary as #1 on the list"). One possible account would involve decomposing *rank* into *cause to come*. On this view, *John ranked [Mary]$_F$ first* has the same syntax as *John caused Mary to come first*. If *first* modifies *come*, we get the list reading, where the order of "coming" matters; if *first* modifies *John cause [Mary]$_F$ to come*, we get the reading where the order of "being caused to come" matters.

[33] $[\![\mathrm{P}]\!]^{t,g}$ is not always $[\![\mathrm{comes}]\!]^{t,g}$ in sentences like (67b), e.g., in the most salient reading of *Obama was before Trump*, $[\![\mathrm{P}]\!]^{t,g} = [\![\mathrm{president}]\!]^{t,g}$.

[34] This account extends to *"a" is first*.

7.1 Alternative decompositions

In Sect. 6, I assumed that *before* and *after* are not internally complex, an assumption that led me to adopt a Bobaljik-style semantics for *-est*. In this subsection, I critique two alternative superlative decompositions of *first*/*last*. The first keeps our Bobaljik-inspired entry for *-est* but assumes that *before* and *after* are suppletive morphological comparatives. On this analysis, *before* and *first* might be *fore* + *-er* and [fore [-er -est]], respectively. This decomposition of *first*/*last* is fully parallel to Bobaljik's decomposition of other superlatives.

The other decompositional analysis I assess draws on Heim's (1999) theory of morphological superlatives. Heim's theory dictates that *first*/*last*, if morphological superlatives, decompose into *-est* + a gradable predicate. *Before* and *after* are not gradable predicates like *tall* or *good*, so we cannot say on Heim's (1999) theory that *first* = *before* + *-est* and *last* = *after* + *-est*. In fact, the only way in Heim's system to encode a morphological relation between *first* and *before* and between *last* and *after* is to say that *before*/*first* and *after*/*last* are comparative-superlative pairs for the same positive. For example, *before* and *first* might be *fore*+*-er* and *fore*+*-est*.

These two variants of my analysis suffer from the fatal flaw that they—unlike the version of my analysis presented in Sect. 6—are powerless to explain the contrast between *first*/*last* and other superlatives regarding "upstairs *de dicto* readings" (Sect. 4), repeated in (68).

(68) Context: John thought: "I want to take some train or other between 1pm and 2pm." Bill thought: "I want take some train or other between 3pm and 4pm." Mary thought: "I want to take some train or other between 5pm and 6pm." None of these people know anything about each other.

a. John wanted to take the earliest train.

b. ??/*John wanted to take the first train.

c. ??/*Mary wanted to take the last train.

Let us start by looking at the predictions of the two non-preferred variants of my analysis, beginning with the variant where *earliest*, *first*, and *last* all decompose into a gradable predicate + Heim's *-est* (call it *-est*^{H}). Heim's (1999) semantics for superlatives is well-equipped to handle the upstairs *de dicto* reading of (68a). On a Heimian account of (68a), the upstairs *de dicto* reading arises when focus is on the matrix subject and *-est*^{H} moves outside the embedded clause, as in (69). Those who use Heim's semantics for superlatives generally agree that the upstairs *de dicto* reading is only derivable by *-est*^{H}-movement (see Charnavel 2023 for recent discussion). When looking at (69), recall that *-est*^{H} has the same meaning as [-er -est] from Sect. 6.1.2. Thus, C in (69) is constrained to be a subset of (70a), and (69) expresses the truth conditions in (70b). While I gloss over the compositional interpretation of [-est^H C]'s sister for simplicity, note that following von Fintel and Heim (1997–2023), I assume that material above and below *want* gets interpreted *de re* and *de dicto*, respectively.[35]

(69) [-est^H C][∼C [λd. [[John]$_F$ [want [PRO to take a d-early train]]]]]

[35] Like Heim (1999), I ignore tense in (69).

(70) a. ALT-SEM([λd. [[John]$_F$ [want [PRO to take a d-early train]]]]) =
{[λd.∀w′ compatible with John's wants: John takes a d-early train in w′],
[λd.∀w′ compatible with Bill's wants: Bill takes a d-early train in w′],
[λd.∀w′ compatible with Mary's wants: Mary takes a d-early train in w′],
. . . }

b. ⟦(69)⟧ = 1 iff
∀Q [[Q ∈ C and Q ≠ [λd.∀w′ compatible with John's wants: John takes a d-early train in w′]] →
MAX([λd.∀w′ compatible with John's wants: John takes a d-early train in w′]) > MAX(Q)]

Put informally, (69) is true iff there is a degree of earliness d such that (a) for each of John's desire worlds w', there is some d-early train in w' that John takes; and (b) for every relevant individual y other than John, y does not take a d-early train in at least one of their desire-worlds (either because y takes no trains in that desire-world or because the train(s) they take in that desire-world are later than d). These truth conditions capture the upstairs *de dicto* reading of (68a), i.e., the reading where *d-early train* is read *de dicto* but the comparison between John and other relevant individuals is *de re*.

While a Heim-style decomposition for *earliest* accounts for (68a), a Heim-style decomposition of *first*/*last* would make incorrect predictions for (68b-c). We can demonstrate the problem via the following informal reasoning for *first*. On the theory in question, *first* decomposes into Heim's *-est* + a gradable predicate (let's suppose it's *fore*).[36] To capture the synonymy of *the earliest train* and *the first train*, Heim would need to use an entry for *fore* that is akin to *early*. Given all this, a Heim-style decomposition of *first* necessarily predicts (71) to be an available LF for (68b), an LF which—if *fore* is similar to *early*—gives us an upstairs *de dicto* reading for (68b) similar to the reading predicted for (68a). Even if *fore* is written to differ from *early* somehow, (71) will still derive some sort of upstairs *de dicto* reading. As such, a Heim-style decompositional analysis of *first* (and *last*) is powerless to avoid a bad prediction about (68b) (and (68c), for similar reasons).

(71) [-est^H C][∼C [λd. [[John]$_F$ [want [PRO to take a d-fore train]]]]]

The contrast in (68) also constitutes an obstinate problem for an analysis that uses a Bobaljik-style semantics for superlatives but decomposes *earliest* and *first*/*last* in a parallel way. To see this, suppose that *earliest* = [early [-er -est]] and *first* = [fore [-er -est]]. This theory can derive the upstairs *de dicto* reading of (68a) if, for instance, we assume the analysis of branching affix structures developed in Sect. 6.1.2: on this analysis, [-er -est] has the same meaning as *-est*^{H} and can scope in the way shown in (69). The issue is that the [-er -est] subconstituent of [fore [-er -est]] can presumably scope in just the same way as the [-er -est] subconstituent of [early [-er -est]], deriving some sort of upstairs *de dicto* reading for *first* as well as *earliest*.

[36] The following reasoning also applies to rule out a theory where *first* and *last* are suppletive superlatives of *early* and *late* (see fns. 4 and 14)—the judgments on (68a) vs. (68b) and the judgments on (68c) vs. *Mary wants to take the latest train* are inexplicable on such a theory.

As just shown, the non-preferred decompositions struggle in the face of (68) because they decompose *first*/*last* and *earliest* in a parallel way. Section 6's analysis, by contrast, can start to make sense of (68) because it posits a distinct internal structure for *first* and *last* vis-à-vis other superlatives: *first* (*before-est*) and *last* (*after-est*) do not contain a gradable adjective, but other superlatives (e.g., *earliest*, [early [-er -est]]) do.

Armed with this difference between *first*/*last* and other superlatives, Sect. 6's analysis can account for (68). The availability of an upstairs *de dicto* reading for (68a) follows straightforwardly for the reason mentioned above: on Sect. 6's analysis, the [-er -est] subconstituent of [early [-er -est]] has the same meaning as *-est*^{H} and can take scope in the way shown in (69). We can rule out an upstairs *de dicto* reading for (68b) because we cannot derive such a reading via movement of *before-est* or *-est*.[37] Starting with *before-est*, consider the structure if *before-est* moved:

(72) [[before -est] C][∼C [λ**x.** ... [DET [**x** [train P]]]]]]]

As it turns out, the derivation crashes regardless of the type of the trace. To start, note that there will be type mismatch between [[before -est] C] (type ⟨dt,t⟩) and its sister if the trace is of a type other than i, d, ⟨dt,t⟩, or ⟨it,t⟩ (recall that I treat i as a subtype of d). Let us consider each of these possibilities in turn. First, if the trace is of type *d*, there is clear mismatch between the trace and its sister $[\![\text{train P}]\!]^{t,g}$, which is ⟨e,t⟩. Second, if the trace is of type *i*, it cannot combine with $[\![[\text{train P}]]\!]^{t,g}$ either, even by IFA. $[\![\beta]\!]^{t,g}$ and $[\![\gamma]\!]^{t,g}$ can only combine via IFA if $[\![\beta]\!]^{t,g}$ is a function that can apply to its sister's intension (see fn. 22). The type-i trace and [train P] cannot serve as β and γ or vice versa. Note that the issues with type-i traces and type-d traces in configurations like (72) are highly general and would persist even if one tinkered with, e.g., the position of the trace within its DP. Finally, if the trace is of type ⟨dt,t⟩ or ⟨it,t⟩, the derivation would presumably involve semantic reconstruction (even if one found a way to make the types work out downstairs), and thus I find it highly unlikely in principle that such an LF could derive an upstairs *de dicto* reading. Even if we discovered an upstairs *de dicto* derivation that involves a trace of type ⟨dt,t⟩ or ⟨it,t⟩, however, one could block this derivation by appealing to a ban on traces of functional types (Poole 2024, a.o.).

Next, consider an *-est*-movement structure.

(73) -est [λ**x**...]

-est, as defined in (40c), needs a first argument of type ⟨dt,⟨dt,t⟩⟩ or ⟨it,⟨it,t⟩⟩, and thus its derived sister in (73) would need to have one of these types to avoid crash.[38] The only situation where *-est*'s derived sister in (73) would be of type ⟨dt,⟨dt,t⟩⟩ or ⟨it,⟨it,t⟩⟩ is a situation where (a) **x** is of type ⟨d,t⟩ or ⟨i,t⟩; (b) λ**x** abstracts over a ⟨dt,t⟩ or ⟨it,t⟩ node. As a result, derivations with *-est*-movement are only possible in my system for LFs with a ⟨dt,t⟩ or ⟨it,t⟩ node for *-est* to move to. In the sort of LF

[37] I focus on (68b) rather than (68c) in the rest of this subsection for expository simplicity, but it should be clear that the predictions about (68c) are the same as the predictions for (68b).

[38] Technically, *-est*'s derived sister could take it as an argument, but this would result in semantic reconstruction and thus not derive an upstairs *de dicto* reading.

for (71) I have been assuming, there is no ⟨dt,t⟩ or ⟨it,t⟩ node for *-est* to move to, and hence a derivation with *-est*-movement is blocked. Even if one somehow came up with an LF for (71) that contains a suitable target for *-est*-movement, one would still need to find a way to give *-est* its remaining arguments; even if one did that (and derived an upstairs *de dicto* reading in the process), any derivation with *-est*-movement that makes a bad prediction could still be blocked by appealing to a ban on traces of functional types.

There are aspects of this analysis that need further scrutiny. For instance, the explanation for why *before-est* movement is blocked relies on IFA, a rule that not all theories adopt. Regardless, I hope to have shown that Sect. 6's decomposition—unlike alternative decompositions—can gain traction on the puzzle of (68a-c). Since the contrast between (68a) and (68b-c) is *prima facie* problematic for a decompositional superlative analysis of *first*/*last*, the fact that Sect. 6's decomposition is able to capture this contrast speaks in its favor vis-à-vis alternatives.

At this point, one might object that Sect. 6's analysis suffers from its own overgeneration issue that alternatives do not, an issue related to *-est*'s selectional restrictions. On Sect. 6's analysis, *-est* wants a first argument of type ⟨dt,⟨dt,t⟩⟩, and while ⟨dt,⟨dt,t⟩⟩ functions are a closed class, there are presumably ⟨dt,⟨dt,t⟩⟩ functions other than *-er*, *before*, and *after*. For example, perhaps the equative head *as* is ⟨dt,⟨dt,t⟩⟩ (74). Without further constraints, Sect. 6's analysis predicts that we should have adjectives like *[tall [as -est]] (*as tall as everyone else*) alongside [tall [-er -est]].

(74) $[\![\text{as}]\!] = \lambda D'.\ \lambda D.\ \text{MAX}(D) \geq \text{MAX}(D')$

A traditional Bobaljik (2012)-style analysis avoids this prediction by positing that *-est* selects for *-er* only. I need a weaker set of selectional restrictions, however, as I crucially need my *-est* to attach to not just *-er* but also *before* and *after* (which I do not treat as morphological comparatives). Hence, Sect. 6's analysis faces the unique problem of finding a way to weaken *-est*'s selectional restrictions while avoiding overgeneration.

I claim that while the non-preferred decompositions cannot gain any traction regarding their upstairs *de dicto* problem (see above), Sect. 6's analysis can gain traction on its potential overgeneration issue. To substantiate this claim, I mention one possible path towards reformulating *-est*'s selectional restrictions (though I leave a fuller solution to future research): we can say that *-est*, as defined in Sect. 6, selects for functions that express degree-based or time-based precedence ($<$ or $>$). *Before*, *after*, and *-er* have this property, but other potential ⟨dt,⟨dt,t⟩⟩ functions (e.g., *as*, *since*) do not.

To conclude this subsection, I note that there may be a reason other than the upstairs *de dicto* facts to doubt variants of my analysis that treat *before*/*after* as comparatives with *-er*: namely, the existence of empirical differences between *before*/*after* and morphological comparatives. To take two examples (among many, see Penka and von Stechow 2011 and Overfelt 2021), *after* differs from morphological comparatives in NPI licensing, as shown in (75); both *before* and *after* differ from morphological comparatives in the licensing of gapping.

(75) a. Caleb arrived later than anyone else did.
b. *Caleb arrived after anyone else did.

(76) a. Tom read the article earlier than Sam the book.
b. *Tom read the article before/after Sam the book.

The differences between *before*/*after* and comparatives given in (75)-(76) are unexpected under a theory where *before*/*after* and comparatives all decompose into *-er* + some gradable adjective. The contrast in (76) is particularly problematic under such a theory: the licensing conditions for gapping are generally thought to require a particular type of coordination structure, so (76) indicates that *before*/*after*-sentences have a structure distinct from that of comparatives (Overfelt 2021).

While (75)-(76) provide suggestive evidence against the non-preferred decompositions discussed at the outset of this section, I leave to future work the question of whether these data furnish an argument against these decompositions that is on a par with the argument from upstairs *de dicto* readings.

7.2 An analysis based on Coppock (2016)

Having critiqued alternative decompositions of *first* and *last*, I next consider a theory where we still treat *first* as *before-est* and *last* as *after-est* but formalize this decomposition using Coppock's (2016) implementation of the Containment Hypothesis rather than the one developed in Sect. 6.1.2. For Coppock (2016), superlatives like *tallest* have the "nesting" structure [[tall -er] -est]. In *tallest student*, *-est* (77) takes the phrasal comparative *taller* (78b), takes *student*, and returns the singleton containing the student who bears the *taller* relation to all others.

(77) $[\![\text{-est}]\!] = \lambda R_{\langle e,et\rangle}.\ \lambda C_{\langle e,t\rangle}.\ \lambda x.\ \forall y\ [[y \in C \text{ and } y \neq x] \rightarrow R(y)(x) = 1]$

(78) a. $[\![\text{tallest student}]\!] = [\![[[[\text{tall -er}^{phrasal}]\text{ -est}]\text{ student}]]\!]$
b. $[\![\text{tall-er}^{phrasal}]\!] = \lambda y.\ \lambda x.\ \text{MAX}(\lambda d.\ x\text{'s height} \geq d) > \text{MAX}(\lambda d.\ y\text{'s height} \geq d)$
c. $[\![(78a)]\!] = \lambda x.\ \forall y\ [[y \text{ is a student and } y \neq x] \rightarrow \text{MAX}(\lambda d.\ x\text{'s height} \geq d) > (\lambda d.\ y\text{'s height} \geq d)$

One could imagine pursuing an analysis where *first*/*last* = *before*/*after* + Coppock's (2016) *-est*, but I opted for the formalization in Sect. 6 instead for three reasons. First, using (77) presupposes that *-er* is ambiguous between a clausal entry and a phrasal entry, and extending Coppock's (2016) analysis to *before-est* and *after-est* would necessitate positing a similar phrasal/clausal ambiguity for *before* and *after*. While some believe that *-er*, *before*/*after*, or both are ambiguous in this way, many others believe that the only entries for these expressions are clausal (see Lechner 2020 on comparatives; Penka and von Stechow 2011 and Overfelt 2021 on *before*/*after*). Thus, the formalization in Sect. 6—which only uses clausal entries—is more theory-neutral than a formalization that uses (77).

Second, the analysis of *earliest*, *best*, etc. in Sect. 6 is functionally identical to Heim's (1999) standard analysis and thus inherits all the positive predictions of that analysis (e.g., those discussed in Howard 2014 and Charnavel 2023). The extent to which a theory with (77) can mimic Heim's (1999) predictions is not clear, as Coppock (2016) introduces (77) in the context of a theory of "superlative modifiers" like *at least* and not in the context of a full-fledged theory of superlatives.

Finally, a theory with *before-est* and *after-est* that uses (77) faces a threat of overgeneration that is bigger than the parallel threat faced by Sect. 6's analysis. As discussed in Sect. 7.1, Sect. 6's analysis faces the question of how to reformulate *-est*'s selectional restrictions so that *-est* can attach to *-er*/*before*/*after* but not other ⟨dt,⟨dt,t⟩⟩ functions. A formalization of my main claim that uses (77) would face the parallel question of how to reformulate *-est*'s selectional restrictions so that *-est* can attach to (phrasal) *-er*/*before*/*after* but not other ⟨e,et⟩ functions. Since ⟨e,et⟩ functions, unlike ⟨dt,dtt⟩ functions, are an open class, the question of about *-est*'s selectional restrictions is more urgent because there is a mind-boggling number of words that can be the sister of (77) without further constraints. This long list of potential sisters must be pared down: otherwise, we predict the existence of an adjective like **brother-est* that picks out the people in C who are brothers of everyone else in C.

(79) a. $[\![\text{brother}]\!] = \lambda y.\ \lambda x.\ x \text{ is the brother of } y$

b. $[\![\text{John is the [[brother-est] C}]\!]$, when defined, $= 1$ iff $\text{John} = \iota x.\ \forall y\ [[y \in C \text{ and } y \neq x] \rightarrow x \text{ is the brother of } y]$

Though the overgeneration threat faced by a *before-est*/*after-est* analysis with (77) may be solvable, the overgeneration threat faced by Sect. 6's *before-est*/*after-est* analysis is much smaller from the get-go, which gives it an advantage.

7.3 A lexicalist alternative

Having zeroed in on the preferred decompositional analysis of *first*/*last*, I briefly compare it to the lexicalist analysis mentioned in Sect. 5, an analysis where *first*/*last* are monomorphemic but have meanings identical to what we would get if we decomposed them as *before-est*/*after-est*.

One could argue in favor of Sect. 6's analysis and against this lexicalist alternative if we could establish that *-est* scopes independently of *before*/*after*; however, the upstairs *de dicto* facts suggest that there is no scopally active subconstituent of *first* and *last*. Without a scope argument for Sect. 6's decomposition, it is very difficult to adjudicate between a lexicalist analysis and a decompositional analysis where the scopal inertness of *-est* in *first* and *last* follows from independent principles (such as the analysis provided in this paper, which can derive the scopal inertness of *-est* in *first*/*last* by appealing to a ban on functional traces). Here is one tentative suggestion about how one might adjudicate between the two theories: we could construct an argument one way or the other if we could find a word that we know requires *-est* as an argument. One candidate for such a word is actually *n-th*, which—according to Alstott (2023) (see Sect. 2.2)—is only licensed next to an overt or covert *-est*. *N-th* can co-occur with *first*/*last* (see Sect. 4), which would suggest that they contain *-est* if we adopt all of Alstott's (2023) assumptions. Unfortunately, there is not enough literature on Alstott's (2023) data to conclude that this is the only possible account for said data, so the argument does not yet go through. I leave further consideration of the lexicalist analysis to future research.

8 Why aren't *first* and *last* ordinals?

Having argued for and formalized my main claim about *first*/*last*, I zoom out and address the bigger-picture question of why the English *first*/*last* are not ordinals. After all, there are languages whose terms for 'first' transparently derive from 'one' + ordinal morpheme (such as Mandarin; see Sect. 4), so we can easily imagine a world where English *first* = **one-th*. Similarly, we can easily imagine a world where *last* is a non-decomposed item with an ordinal syntax-semantics. Why are these worlds not the actual world?

The only previous scholar to address this question is Barbiers (2007), who focuses on the Dutch superlative *eerst* 'first' (see Sect. 3). He argues that the Dutch equivalent of **one-th* is blocked because the ordinal morpheme selects for numerals that (when predicated of a noun) yield a multi-membered set that can be non-vacuously ordered. *Two*, *three*, etc. yield sets that can be non-vacuously ordered when predicated of a noun, but *one* yields a singleton set—and singleton sets cannot be meaningfully ordered. The ill-formedness of **one-th* leaves a meaning gap, Barbiers (2007) says, and so Dutch uses a superlative to approximate the meaning of **one-th*.

Building on and formalizing this account, I suggest that there are two types of ordinal morphemes cross-linguistically: ordinal morphemes that yield a well-formed meaning when attached to 'one', and ordinal morphemes that—because of the special properties of 'one' noted by Barbiers (2007)—do not. More concretely, we can assume that languages like Mandarin have the ordinal morpheme in (3a) above, while English *-th* is akin to (80a). For simplicity, I assume that o(P) is a strict total order.[39]

(80) a. $[\![\text{n-th}]\!] = \lambda C_{\langle e,t\rangle}.\ \lambda P_{\langle e,t\rangle}.\ \lambda x\text{: } x \in C \cap P \text{ and } |C \cap P| \geq n.$
$\exists!Q\text{: } Q \subseteq C \cap P \wedge |Q| = n \wedge x \in Q \wedge \forall y \in Q\ [y \neq x \rightarrow y >_{o(P)} x]$

b. $[\![[\text{n-th C}]\ \text{train}]\!] = \lambda x\text{: } x \in C \cap [\![\text{train}]\!] \text{ and } |C \cap [\![\text{train}]\!]| \geq n.$
$\exists!Q\text{: } Q \subseteq C \cap [\![\text{train}]\!] \wedge |Q| = n \wedge x \in Q \wedge \forall y \in Q\ [y \neq x \rightarrow y >_{o(C \cap [\![\text{train}]\!])} x]$

(3a) yields a well-formed meaning for all numerals including 1, and thus Mandarin has *one-th*. By contrast, (80a) yields a well-formed meaning **except** when n = 1. To see this, suppose that there are three trains (A, B, and C) such that $A >_{o(C \cap [\![\text{train}]\!]))} B >_{o(C \cap [\![\text{train}]\!])} C$; let us compare the predicted meanings of *second train* and *one-th train* with (80a). $[\![\text{second train}]\!]^{t,g}$ is true of a train $x \in \{A, B, C\}$ iff there is a unique doubleton subset of {A, B, C} containing x in which x is last. B is the only member

[39] If one adopts the idea that there is a covert *first* in *the second train* in English (see Sect. 5), one can use (i) instead.

(i) $[\![\text{n-th}]\!] = \lambda \text{SUP}_{\langle dtt,dtt\rangle}.\ \lambda C_{\langle dt,t\rangle}.\ \lambda D.\ \exists!Q\text{: } Q \subseteq C \wedge |Q| = n \wedge D \in Q \wedge \forall D' \in Q\ [D \neq D' \rightarrow \text{SUP}(\{D,D'\})(D') = 1]$

I leave it to the reader to verify that this nets the correct meaning for *second ~~first~~ train* (LF: [λx. [[[before -est] two-th] C][~C [[x]$_F$ train P]]]) and rules out **one-th* in the same way as (80a). Note that even if one adopts (i) for English, there are likely other languages that use entries closer to (3a) or (80a): ordinals cannot modify overt superlatives in languages like Mandarin (Yi-Hsun Chen, p.c.), and standard Dutch (Ruby Sleeman, p.c.), so it is implausible to give these languages an entry for *n-th* that requires *-est* as an argument.

of {A, B, C} with this property—there are no doubleton subsets in which A is last and multiple doubleton subsets in which C is last ({A, C}, {B, C}). *The second train* is thus predicted to denote B (as desired), and we see that *second* has a non-trivial contribution: it picks out B from the set {A, B, C}.

$[\![\text{one-th train}]\!]^{t,g}$ is predicted to be true of a train $x \in$ {A, B, C} iff there is a unique singleton subset of {A, B, C} containing x (note that when $n = 1$, the final conjunct of the right-hand side of (80a) is vacuously satisfied). This property trivially holds of A, B, and C, and hence $[\![\text{one-th train}]\!]^{t,g} = [\![\text{train}]\!]^{t,g}$. Abstracting away from this particular example, $[\![\text{one-th}]\!]^{t,g}$ ends up being essentially an identity function, and so it can be ruled out on grounds of redundancy/triviality.[40] The ill-formedness of **one-th* leaves a meaning gap in English that gets filled by a superlative.[41]

The preceding discussion has focused on *first* rather than *last* because it is hard to glean from surface morphological evidence whether some languages express *last* via an ordinal. After all, ordinal theories of *last* are non-decompositional: they treat *last* as a non-decomposed item with an ordinal syntax-semantics and not as an instance of *n-th* (Bylinina et al. 2014). So while we can establish that some languages express 'first' via an ordinal by looking at morphology, it is difficult to imagine a term for 'last' whose surface form alone forces us to say that it is non-decomposed and has an ordinal syntax-semantics. The only way to establish the status of a language's 'last' is to undertake a syntactic-semantic investigation like the one from Sects. 4-5. Without such investigations for languages besides English, one cannot fruitfully theorize about why English *last* is not an ordinal (since we don't know if this is an English-specific quirk). But since languages differ in whether their 'first' is a superlative or an ordinal, my hunch is that languages will show similar differences in their terms for 'last'.

9 Conclusion

I have argued that *first* and *last* are not ordinals but rather superlative counterparts of *before* and *after*. Focusing on decompositional implementations of this claim, I considered a formalization that uses Heim's (1999) semantics for superlatives and ultimately rejected it in favor of a formalization that uses a Bobaljik (2012)-inspired semantics. Important issues for future research include (a) adjudicating between the Bobaljik (2012)-style decompositional analysis and a lexicalist alternative and (b) solidifying a typology of the semantics of *first*/*last* cross-linguistically.

[40] One can also rule it out by giving (80a) a presupposition that is unsatisfiable whenever n = 1. For example: recalling our assumption that o(P) is a strict order, one could say that [[n-th C] P] is defined only if for every subset of C ∩ P of cardinality n, o(P) relates two members of that subset. A strict order never relates the member of a singleton to itself, so this presupposition is unsatisfiable when $n = 1$. Since (80a)'s assertion quantifies over subsets of cardinality n, this presupposition is fairly natural because it rules out vacuous comparison in such subsets. Similar anti-vacuity presuppositions (e.g., |C| > 1) are often posited for superlatives.

[41] I predict that *one-th* being ill-formed is sufficient but not necessary for the emergence of non-ordinal terms that mean 'first'. Since suppletive and non-suppletive terms for 'first' co-exist in many languages (Barbiers 2007), this prediction is likely borne out (though further investigation is needed).

Acknowledgements I thank Amir Anvari, Cleo Condoravdi, Kai von Fintel, Danny Fox, Martin Hackl, Sabine Iatridou, Viola Schmitt, Clemens Mayr, three anonymous NLS reviewers, and audiences at CreteLing 2023, NELS 54, and MIT LingLunch for their insights and feedback. All remaining errors are my own.

Funding Open Access funding provided by the MIT Libraries.

Declarations

Compliance with ethical standards The novel data reported in this article consist of the author's own linguistic judgments (for English) and the linguistic judgments of several colleagues (for other languages). This article reports no research involving animals or non-linguist human participants.

Competing Interests No funds, grants, or other support was received. The author has no competing interests to declare that are relevant to the content of this article.

References

Alstott, Johanna. To appear. On very-intensified superlatives. *Proceedings of Sinn und Bedeutung (SuB)* 29.

Alstott, Johanna. 2023. Ordinal numbers: Not superlatives, but modifiers of superlatives. In *Proceedings of Semantics and Linguistic Theory (SALT) 33*, eds. Juhyae Kim, et al., 1–20. https://doi.org/10.3765/snxdqy39.

Anscombe, Elizabeth. 1964. *Before* and *after*. *Philosophical Review* 73(1):3–24. https://doi.org/10.2307/2183199.

Barbiers, Sjef. 2007. Indefinite numerals one and many and the cause of ordinal suppletion. *Lingua* 117:859–880. https://doi.org/10.1016/j.lingua.2006.03.003.

Beaver, David, and Cleo Condoravdi. 2003. A uniform analysis of *before* and *after*. In *Proceedings of Semantics and Linguistic Theory (SALT) 13*, eds. Robert. B. Young and Yuping Zhou, 37–54. https://doi.org/10.3765/salt.v13i0.2899.

Bhatt, Rajesh. 2006. *Covert modality in non-finite contexts*. Berlin: De Gruyter Mouton.

Bobaljik, Jonathan. 2012. *Universals in comparative morphology: Suppletion, superlatives, and the structure of words*. Cambridge: MIT Press.

Bumford, Dylan, and Yael Sharvit. 2022. Negative polarity items in definite superlatives. *Linguistic Inquiry* 53(2):255–293. https://doi.org/10.1162/ling_a_00409.

Bylinina, Lisa, Natalia Ivlieva, Alexander Podobryaev, and Yasutada Sudo. 2014. A non-superlative semantics for ordinals and the syntax and semantics of comparison classes. Manuscript, Meertens Institut, Institut Jean-Nicod, Higher School of Economics, and UCL.

Charnavel, Isabelle. 2023. Appealing to superlative clauses. Manuscript, University of Geneva.

Condoravdi, Cleo. 2010. NPI licensing in temporal clauses. *Natural Language and Linguistic Theory* 28:877–910. https://doi.org/10.1007/s11049-010-9115-z.

Coppock, Elizabeth. 2016. Superlative modifiers as modified superlatives. In *Proceedings of Semantics and Linguistic Theory (SALT) 26*, eds. Mary Moroney, et al., 471–488. https://doi.org/10.3765/salt.v26i0.3822.

Davies, Mark. 2008. American English (COCA). In *The corpus of contemporary*. https://english-corpora.org/coca/.

del Prete, Fabio. 2008. A non-uniform semantic analysis of the Italian temporal connectives prima and dopo. *Natural Language Semantics* 28:157–203. https://doi.org/10.1007/s11050-008-9030-6.
Fitzgibbons, Natalia, Yael Sharvit, and Jon Gajewski. 2008. Plural superlatives and distributivity. In *Proceedings of Semantics and Linguistic Theory (SALT)*, eds. Tova Friedman and Satoshi Ito. Vol. 18, 302–318. https://doi.org/10.3765/salt.v18i0.2501.
Fox, Danny, and Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19:87–107. https://doi.org/10.1007/s11050-010-9065-3.
Heim, Irene. 1999. Notes on superlatives. Manuscript, MIT.
Heim, Irene. 2000. Degree operators and scope. In *Proceedings of Semantics and Linguistic Theory (SALT) 10*, eds. Brendan Jackson and Tanya Matthews, 40–64. https://doi.org/10.3765/salt.v10i0.3102.
Heinamäki, Orvokki. 1974. *Semantics of English temporal connectives*, PhD dissertation, University of Indiana.
Herdan, Simona, and Yael Sharvit. 2006. Definite and nondefinite superlatives and NPI licensing. *Syntax* 9(1):1–31. https://doi.org/10.1111/j.1467-9612.2006.00082.x.
Heycock, Caroline. 2005. On the interaction of adjectival modifiers and relative clauses. *Natural Language Semantics* 13:359–382. https://doi.org/10.1007/s11050-005-4533-x.
Howard, Edwin. 2014. Superlative degree clauses: Evidence from NPI licensing, MA thesis, MIT.
Krifka, Manfred. 2010. Before and after without coercion: Comment on the paper by Cleo Condoravdi. *Natural Language and Linguistic Theory* 28:911–929. https://doi.org/10.1007/s11049-010-9116-y.
Larson, Richard. 2000. ACD in AP? Paper presented at the West Coast Conference on Formal Linguistics (WCCFL) 19.
Lechner, Winfried. 2020. Clausal vs Phrasal Comparatives, 1–39. New York: Wiley.
Levin, Yael. 2017. Who hobbles after the subject: Parables of writing in *The Third Policeman and Molloy*. *Journal of Modern Literature* 40(4):105–121. https://doi.org/10.2979/jmodelite.40.4.08.
Overfelt, Jason. 2021. Stripping and VP ellipsis in reduced temporal adverbs. *Syntax* 24(4):462–509. https://doi.org/10.1111/synt.12222.
Partee, Barbara. 1986. Noun phrase interpretation and type-shifting principles. In *Studies in discourse representation theory and the theory of generalized quantifiers*, eds. Jeroen Groenendijk, Dick de Jongh, and Martin Stokhof, 115–143. Berlin: de Gruyter. https://doi.org/10.1515/9783112420027-006.
Penka, Doris, and Arnim von Stechow. 2011. Phrasal complements of *before* and *after*. In *Empirical issues in syntax and semantics 8*, eds. Olivier Bonami and Patricia Cabredo Hofherr, 435–451.
Poole, Ethan. 2024. (Im)possible traces. *Linguistic Inquiry* 55(2):287–326. https://doi.org/10.1162/ling_a_00467.
Romero, Maribel. 2013. Modal superlatives: A compositional analysis. *Natural Language Semantics* 21:79–110. https://doi.org/10.1007/s11050-012-9090-5.
Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1:75–116. https://doi.org/10.1007/BF02342617.
Rudnitski, Rose. 1995. Patty Smith Hill, gifted early childhood educator of the progressive era. *Roeper Review* 18:19–24. https://doi.org/10.1080/02783199509553691.
Schwarz, Bernhard. 2005. Modal superlatives. In *Proceedings of Semantics and Linguistic Theory (SALT) 15*, eds. Efthymia Georgala and Jonathan Howell, 187–204. https://doi.org/10.3765/salt.v15i0.3094.
Sharvit, Yael. 2010. Infinitival superlatives: English vs. Modern Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics* 2(1):213–247. https://doi.org/10.1163/187666310X12688137960821.
Sharvit, Yael. 2015. The onliest NP: Non-definite definites. In *West coast conference on formal linguistics (WCCFL) 32*, eds. Ulrike Steindl, et al., 171–190.
Stateva, Penka. 2002. *How different are different degree constructions?* PhD dissertation, University of Connecticut.
Szabolcsi, Anna. 1986. Comparative superlatives. In *MIT working papers in linguistics 8*, eds. Naoki Fukui, Tova Rapoport, and Elizabeth Sagey, 245–265.
Szabolcsi, Anna. 2012. Compositionality without word boundaries: (the) more and (the) most. In *Proceedings of Semantics and Linguistic Theory (SALT) 22*, ed. Anca Chereches, 1–25. https://doi.org/10.3765/salt.v22i0.2629.
von Fintel, Kai, and Irene Heim. 1997–2023. Intensional semantics. Manuscript, MIT.
Yee, Charles. 2010. Building DRT lexical entries for superlatives and ordinal numbers. Manuscript. University of Stuttgart.