The source of nonfinite temporal interpretation*

Ellise Moon & Aaron Steven White

University of Rochester

1. Introduction

Which aspects of semantic interpretation are due to predicates' denotations and which are due to the denotations of their arguments? This question has proven particularly difficult to answer in the context of the temporal interpretation of nonfinite embedded clauses (Stowell 1982, Martin 2001, Katz 2001, Abusch 2004, Landau 2000, Wurmbrand 2001, 2014, Grano 2015, Pearson 2016, Williamson 2019). For example, is it a fact about *regret* or its subordinate clause that explains why the leaving referred to in (1) can be yesterday but not tomorrow? And is the answer analogous for the opposite pattern observed with *want* in (2)?

(1) a. Jo will regret leaving yesterday.
 b. #Jo regretted leaving tomorrow.

(2) a. #Jo will want to leave yesterday.
 b. Jo wanted to leave tomorrow.

One reason this question is difficult to answer is that temporal interpretation appears to correlate with the predicate in some cases, but with the embedded clause structure in others. For instance, keeping the predicate constant, a change in embedded clause structure can alter the pattern: in (3), *remember* patterns like *regret*, but in (4), it patterns like *want*.

(3) a. Jo will remember leaving yesterday.
 b. #Jo remembered leaving tomorrow.

(4) a. #Jo will remember to leave yesterday.
 b. ?Jo remembered to leave tomorrow.

In contrast, *propose* patterns like *want* regardless of the embedded clause structure.

(5) a. #Jo will propose leaving yesterday.
 b. Jo proposed leaving tomorrow.

(6) a. #Jo will propose to leave yesterday.
 b. Jo proposed to leave tomorrow.

Is the pattern in (3) and (4) an idiosyncratic fact about *remember* or a general fact about the composition of its complements? And if the latter, how do we explain (5) and (6)?

*We would like to thank audiences at NELS 50 and the University of Rochester for useful comments on this work. This work was funded by an NSF grant: *The MegaAttitude Project: Investigating selection and polysemy at the scale of the lexicon* (BCS-1748969/BCS-1749025).

We approach such questions about *temporal orientation*—i.e. how an eventuality expressed in an embedded clause is temporally located relative to the one described in that clause's matrix—by (i) using an acceptability judgment task to collect a large sample of patterns like those in (1)-(6), covering a broad swath of the lexicon (§3); and (ii) developing a computational model for extracting systematicities in those patterns across lexical items and embedded clause structures (§4).[1] We set our model up so as to be general across a range of possible accounts proposed in prior work (§2), with each setting of our model's hyperparameters corresponding to a different instantiation of some such account. We then use standard quantitative model comparison techniques to determine which hypotheses best explain observed data, while simultaneously predicting the patterns in unseen data well. In this comparison, we find that models attributing temporal orientation only to predicate denotations—not to complement denotations—reliably underperform those that attribute it to only complement denotations or both predicate and complement denotations.

2. Background

Approaches to temporal orientation fall into three broad classes: (i) *lexical* approaches, which posit that temporal orientation is attributable to the predicate, with no contribution from the nonfinite complement; (ii) *structural* approaches, which posit that temporal orientation is attributable to a covert element—usually, a modal or tense—in the nonfinite complement selected by some predicate; and (iii) *mixed* approaches, which posit (a) that some predicates impose contraints on temporal interpretation while others rely on tense/aspect in their complement to do so; and/or (b) that predicates somehow interact with the denotation of their nonfinite complement to impose constraints on temporal interpretation.

Lexical approaches have intuitive appeal for at least two reasons. First, they allow one to take the apparent lack of tense, aspect, or modality morphology in nonfinites at face value: if nonfinite complements do not contain a temporal operator, they must get their temporal interpretation from somewhere else. Second, for many nonfinite-taking predicates, there are often clear intuitions about the content of the eventualities described by those predicates: regrets are about the past, while proposals are about the future, so it seems natural to build temporal orientation into the denotations of predicates like *regret* and *propose*.

Lexical approaches can be implemented in a variety of ways (see Ogihara 1996, Katz 2001, Abusch 2004, Pearson 2016), but as Katz notes, they face a challenge when attempting to explain the interpretation of finite complements: in (7a), the expectation holds before the raining—the sentence is *future-oriented*—and thus, in a lexical approach, the denotation of *expect* would need to enforce that the eventuality described in its complement happens/holds after the expectation (starts); but such a denotation for *expect* incorrectly predicts that the expectation described in (7b) similarly precedes the raining.

(7) a. Jo expects it to rain {#right now, an hour from now}.
   b. Jo expects that it is raining {right now, #an hour from now}.

[1] We diverge from most prior work in referring to sentences (or possibly verb phrases), rather than predicates, as having a particular temporal orientation because, as evidenced by the pattern in (3) and (4), temporal orientation depends not only on the predicate, but also the structure of the clause a predicate combines with.

Indeed, this challenge is general in applying to any case where a change in the structure of the complement results in a change in orientation—the contrast in (3)-(4) being another example. To handle such facts, a lexical account must posit either some form of polysemy for predicates that take both nonfinite and finite complements or some process that effectively neutralizes the temporal orientation that would normally be associated with the predicate.

Alternatively, one might take patterns like that in (7) to suggest that it is wrong to view temporal orientation constraints as imposed directly by the lexical item. Rather, in at least some cases, the embedded clause itself might contain a temporal operator. Such an account might then explain the temporal orientation observed in (7a) by saying that the infinitival contains an operator similar to *will* in (8) (see Bresnan 1972).

(8) Jo expects that it will rain {#right now, an hour from now}.

Stowell (1982) proposes a version of such an account. On his account, which is an early version of a mixed approach, certain nonfinite complements—specifically, infinitival complements in control structures (9)—carry 'unrealized' tense, while others do not—specifically, infinitives in accusative-infinitive (AcI) constructions (10) and raising constructions (11) as well as gerundive complements (12). When a complement does not contain tense, the temporal interpretation is determined by the verb that selects that complement—e.g. *enjoy* in (12a) determines that the winning happened before the enjoying but *try* in (12b) determines that the winning would happen in the future of the trying (if the trying were successful).

(9) a. I expected to win the race.
b. I managed to win the race.

(10) a. I believed Jo to be intelligent.
b. I wanted Jo to win the race.

(11) a. Jo seems to be intelligent.
b. Jo appears to have won the race.

(12) a. I enjoyed winning the race.
b. I tried winning the race.

Martin (2001) proposes a similar account, agreeing with Stowell in positing that control infinitivals contain some element that gives rise to the observed temporal interpretation—for Martin, a modal element similar to *would* or *should* (see also Bhatt 1999)—but differing in positing that gerundive complements also contain such an element (for him a nonmodal 'pure' tense). He suggests that different temporal interpretations are generated by the particular element a particular predicate selects (see also Ogihara 1989).

One quirk of both proposals is that the future orientation of, e.g., (7a) and (9a) comes about by different mechanisms: in (9a), there is some temporal operator in the infinitival that gives rise to future orientation, while in (7a), *expect* itself gives rise to future orientation. This sort of explanation seems to miss the mark in being forced to explain the behavior of *expect* twice. If one is unwilling to take a lexical approach to *expect*—and thereby pay the price of positing polysemy or coercion to explain the pattern in (7)—then another option is to posit that the same operator occurs in the complements of both (7a) and (9a).

Wurmbrand (2014) remedies this by not drawing as tight a connection between aspects of construal and temporal orientation as Stowell and Martin (cf. Pesetsky 1991, Landau 2000, Wurmbrand 2001, Pesetsky and Torrego 2004, Grano 2015). On her account, which focuses on only infinitival (not gerundive complements), predicates select from among three types of nonfinite complements: one containing an anaphoric (or *zero*) tense that is dependent on the embedding predicate for its interpretation, one containing a tenseless future modal, and one containing aspect and no tense. For Wurmbrand, the future infinitival is the one found in both (7a) and (7a). The other two are aimed at explaining *present oriented* (or *simultaneous*) sentences like (10a)—where the time of the believing (or more specifically, argues Wurmbrand, the time at which the believer situates themself) is the same as (or contained by) the time of the being intelligent—and (9b)—where the managing and the winning both occur at the matrix clause reference time.

One challenge that remains for these structural accounts is to explain cases where a complement is apparently too small to host a temporal operator: in (13), it is unclear where the temporal operator would be situated (see discussion in Williamson 2017).

(13) Jo {wanted, expected} the check. $\rightsquigarrow$ Jo {wanted, expected} to have the check.

One way to deal with this is to bite the bullet and say that sentences like (13) do in fact have a covert temporal operator—either as part of the noun phrase or some larger covert structural structure (see Grano 2015, Ch. 4 and refereinces therein). Alternatively, one might posit a mixed approach wherein both the embedding predicate and the nonfinite complement constrain the orientation (Grano 2017, Williamson 2019, see also White 2014). This latter approach would allow the temporal orientation of (13) to arise from the denotation of the predicate while also retaining the benefits of positing a temporal operator within nonfinites. The main downside of such an approach is that it substantially increases the complexity of the account, since for every pairing of a predicate and complement type, one must say whether the temporal orientation is due to the predicate, the complement, or both.

In sum, the prior literature on temporal orientation leaves us with two main question. First, is the most parsimonious account of temporal orientation lexical, structural, or mixed? As discussed above, each type of account has aspects of temporal orientation that it covers parsimoniously and aspects that it does not. Second, assuming some type of account, how many subtypes of predicates and/or nonfinite complements are required to explain the entirety of the nonfinite complement-selecting lexicon?

In the remainder of the paper, we approach these questions quantitatively by (i) using an acceptability judgment task to collect a large sample of temporal orientation patterns, covering a broad swath of the lexicon (§3); and (ii) developing a computational model that is general in the sense that different settings of that model's hyperparameters correspond to different instantiations of a lexical, structural, or mixed approach (§4).

3. The MegaOrientation dataset

To quantitatively measure temporal orientation across a wide range of predicates and nonfinite complements in English, we collect the MegaOrientation dataset (available at megaattitude.io). MegaOrientation contains acceptability judgments for sentences that contain a nonfinite complement and require either a future or a past orientation. We enforce said orientation by manipulating the matrix predicate tense/modality (past v. future) in the sentences as well as a temporal adverb phrase (*in the past, in the future*). This adverbial is paired with an incompatible matrix tense, so as to block a reading of the sentence where the adverbial attaches high. Two example items are shown in (14).

(14) a. Someone remembered doing something in the future.
b. Someone will remember doing something in the past.

The acceptability judgments of the future- and past-oriented items for a particular predicate-complement pair then form our measure of orientation. For instance, because sentences containing *remember* with a gerundive complement are past-oriented, (14a) should receive low acceptability judgments and (14b) should receive high acceptability judgments.

3.1 Predicate-complement selection and item construction

Following White and Rawlins (2018), we select predicate-nonfinite complement pairs from the MegaAttitude dataset (White and Rawlins 2016, in press) based on their acceptability—on average, a 4 or better—for the nonfinite complement types exemplified in (15).[2] We select from MegaAttitude both because it effectively covers the entire English clause-embedding lexicon, and because the *bleaching method* used to construct its items allows us to avoid typicality effects that might add noise to the judgments.

(15) a. Someone wanted to do something. [NP _ to VP[+EV]]
b. Someone seemed to have something. [NP _ to VP[-EV]]
c. Someone was told to do something. [NP _ to VP[+EV]]
d. Someone was believed to have something. [NP _ to VP[-EV]]
e. Someone regretted doing something. [NP _ VPing]

The complement types in (15) are selected so as to span those discussed in the literature described in §2. We include (15a) and (15b) to cover subject control and raising constructions, manipulating the eventivity of the predicate in the complement to ensure that we capture present-oriented predicates (Abusch 2004). We include (15c) and (15d) to cover object control and AcI constructions, using passivization so that both *believe* and *wager* class verbs are covered by the same frame (Pesetsky 1991, Moulton 2009). This selection method yields 887 unique predicates (many occurring in multiple frames) and 2,208 predicate-frame pairs: 513 in [NP _ VPing], and 404 in [NP _ to VP[+EV]], 463 [NP _ to VP[+EV]], 364 in [NP _ to VP[-EV]], and 464 in [NP _ to VP[-EV]].

We construct two sentences for each predicate-complement pair: one with a matrix past tense and embedded future adverb (16a), and another with a future modal in the matrix and an embedded past adverb (16b). This yields 2,208 verb-frame pairs $\times$ 2 orientations = 4,416 unique items.

(16) a. Someone regretted doing something in the future. FUTURE-ORIENTED
b. Someone will regret doing something in the past. PAST-ORIENTED

[2]This selection criterion aims to ensure that any grammatical unacceptability which arises in our dataset is the result of the forced temporal orientation in a given sentence, not ungrammaticality of a given predicate-complement pairing. Of course, this method does not perfectly control for lower acceptability predicate-complement pairings, and so we will attempt to correct for this in our model.

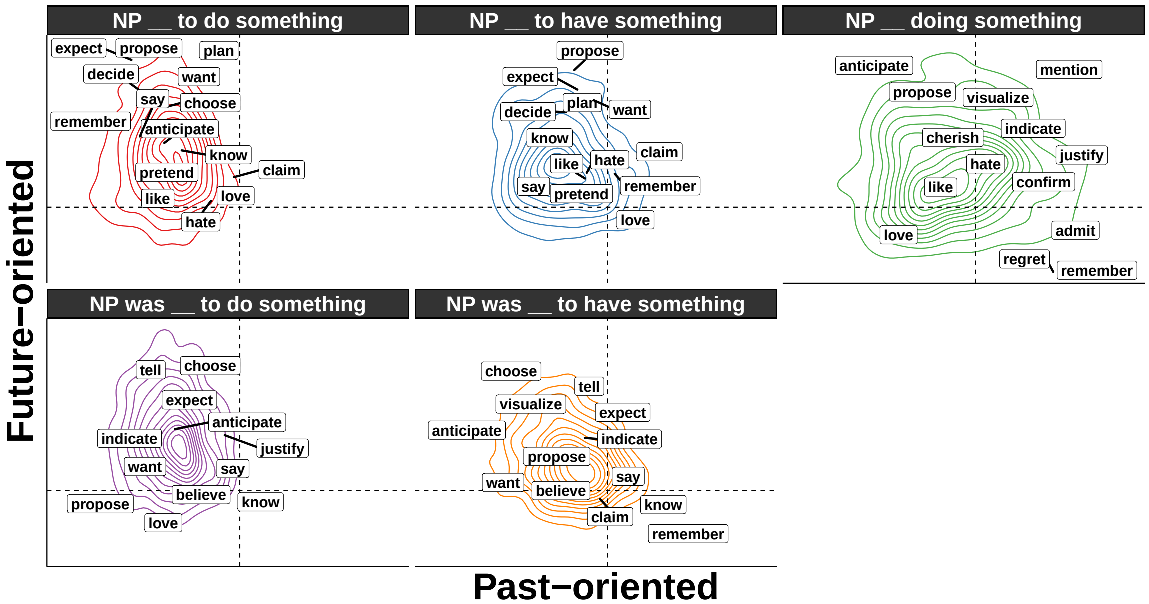

Figure 1: Density plot of normalized acceptability judgement scores for all predicates in the dataset, faceted by embedded clause type. Only a subset of predicates are labeled.

Future interpretation is indicated by high acceptability in (16a) and past interpretation is indicated by high acceptability in (16b). A present orientation is indicated by low acceptability in both. As discussed below, some predicate-complement pairs allow both.

3.2 Methodology

Sentences were divided into lists of 48, with an equal number of future- (16a) and past-oriented (16b) items in each list and as even of a distribution of syntactic frames as possible, with no verb appearing more than once. We follow the method of balancing lists for verb frequency and verb-frame acceptability described by An and White (2020).

Participants were presented with a sentence and were asked to rate the acceptability of the sentences on a 7-point scale, with 1 being *terrible* and 7 being *perfect*. At the beginning of the assignment, instructions were provided, which explained our layperson's definition of *acceptability* as "something a native speaker of English would say, even if the situation the sentence describes sounds vague or implausible." Annotators were also provided with three practice questions before beginning the task to calibrate participants to the scale—one perfectly acceptable, one exteremely unacceptable, and one middling.

3.3 Participants

A total of 869 unique participants were recruited through Amazon's Mechanical Turk to give acceptability judgments. Participants were allowed to do as many lists as they wanted, but no annotator responded to the same list twice. Each list was annotated by 10 distinct annotators. All participants reported being native speakers of English.

3.4 Results

Figure 1 plots the response data for acceptability of predicates in the past-oriented construction (x-axis) against acceptability in the future-oriented one (y-axis) after applying the ordinal model-based normalization procedure described in White and Rawlins in press. Dashed lines show the midpoint of the acceptability scale and are included as a guide to interpreting the plot, rather than a meaningful cutoff.

A predicate being located towards the top left of a plot indicates that it is more future-oriented with that complement type while a predicate towards the bottom right is past-oriented with that complement type. Predicates falling into upper right allow both orientations, and predicates with low acceptability in either orientation are ones with a present-oriented/simultaneous reading with a particular complement type.

The results in Figure 1 suggest that our method correctly captures the orientation of predicate-complement pairs commonly discussed in the literature: *expect*, *want*, and *hope* are future-oriented with infinitival complements; *remember* and *regret* are past-oriented with gerundive complements; and *believe* and *claim* are present-oriented/simultaneous. Further, we see that past orientation is largely confined to predicates with gerundive complements, though a small number of predicates taking infinitival complements appear to be past-oriented, consistent with observations in Karttunen 1971 discussed in Martin 2001 (see Grano 2015, Ch. 5 for a potential explanation of these cases).

4. Model

Our main aim in collecting the data described in §3 is to quantitatively assess the approaches discussed in §2. With this aim in mind, we now describe a computational model whose hyperparameter settings correspond to different instantiations of a lexical, structural, or mixed approach. In conjunction with the data described in §3, this setup allows us to use standard quantitative model comparison techniques to determine which approach best explains observed data, while predicting the patterns in unseen data well.

4.1 White and Rawlins' model of s-selection

Our model is an adaptation of An and White's (2020) model, which is itself an extension of White and Rawlins' (2016) model of s(emantic)-selection. White and Rawlins's model aims to induce predicates' semantic type signatures—e.g. that *think* can denote a relation between an entity and a proposition—from their syntactic distribution—e.g. that *think* is acceptable in NP _ S frames. White and Rawlins formalize this task as a boolean matrix factorization problem: given a boolean matrix $\mathbf{D}$ where $d_{vf} \in \mathbb{B} = \{\top, \bot\}$ and $d_{vf} = \top$ iff predicate v is acceptable in syntactic frame f—e.g. NP _ S—they aim to find boolean matrices $\mathbf{\Lambda}$ and $\mathbf{\Pi}$, where $\lambda_{vt} = \top$ iff verb v has type signature t—e.g. $\langle\langle s,t\rangle,\langle\langle e,t\rangle\rangle\rangle$—$\pi_{tf} = \top$ iff type signature t can project onto frame f, and $d_{vf} \approx \bigvee_t \lambda_{vt} \wedge \pi_{tf}$. Intuitively, a verb is acceptable in a frame if that verb has some type signature that can project onto that frame, and any deviation is considered 'lexical noise', e.g., due to a predicate's c(omplement)-selection properties (Grimshaw 1979) or abstract case (Pesetsky 1982).

White and Rawlins use the MegaAcceptability dataset to obtain a measure of **D**, modeling acceptability judgments for verb v in frame f using an ordinal mixed model with unconstrained random cutpoints for each annotator and $\mathbb{P}(d_{vf})$ as a fixed effect. This allows them to estimate $\mathbb{P}(\lambda_{vt})$ and $\mathbb{P}(\pi_{tf})$ as $\mathbb{P}(d_{vf}) = 1 - \prod_t 1 - \mathbb{P}(\lambda_{vt})\mathbb{P}(\pi_{tf})$ (see their paper for a derivation). They then use standard optimization procedures to obtain $\mathbb{P}(\lambda_{vt})$ and $\mathbb{P}(\pi_{tf})$ under different settings for the number of types assumed.

4.2 Extending White and Rawlins' model with a model of interpretation

White and Rawlins' model provides some basic components for our model, such as the idea that there may be multiple types of structures a predicate can select—e.g. one containing tense and one not (Stowell 1982) or one containing tense, one containing only a tenseless modal, and one not containing tense (Wurmbrand 2014)—and potentially multiple ways for that type to be realized. What it is missing, for our purposes, is (i) a representation of the interpretation of structural types—e.g. that some type requires future orientation while another requires present orientation/simultaneity; and (ii) a representation of lexical types and their interpretations—e.g. that some lexical type does (not) license future orientation.

The basic idea behind our model, which is identical in its mathematical formulation to An and White's, is to add such components while allowing one to independently manipulate whether the model assumes just structural types, just lexical types, or both; and if so, how many of each. The upshot is that, when the model assumes some number of lexical types but no structural types, it corresponds to a lexical approach; when it assumes some number of structural types but no lexical types, it corresponds to a structural approach; and when it assumes some number of lexical items and some number of structural types, it corresponds to a mixed approach. This setup then allows us to use quantitative model comparison to assess which setting of the number of lexical and structural types best fits the observed data, while also explaining held out data well.

Following An and White, we add to White and Rawlins' model: (i) a boolean matrix $\boldsymbol{\Psi}$, wherein $\psi_{vi} = \top$ iff verb $v \in \mathscr{V}$ has lexical type $l \in \mathscr{L}$; (ii) a boolean matrix $\boldsymbol{\Phi}$, wherein $\phi_{lo} = \top$ iff property l licenses orientation $o \in \mathscr{O}$; and (iii) a boolean matrix $\boldsymbol{\Omega}$, wherein $\omega_{to} = \top$ iff semantic type signature $t \in \mathscr{T}$ licenses orientation o. We then model the acceptability of interpreting predicate v in frame f with orientation o as $a_{vfo} \approx \bigvee_{t,l} \lambda_{vt} \wedge \psi_{vl} \wedge \phi_{lo} \wedge \pi_{tf} \wedge \omega_{to}$. Intuitively, a sentence consisting of a verb in some frame can have some temporal orientation if (i) that verb can select some structural type that can be realized by that frame; (ii) that structural type is compatible with that orientation; and (iii) that predicate has some lexical type that is compatible with that orientation.

To capture the possibility that only lexical or only structural types may be active in explaining temporal orientation, we consider two families of what An and White term *boundary models*. In the boundary models that posit no lexical properties, we set $|\mathscr{L}| = 1$ and fix $\mathbb{P}(\psi_{v1}) = 1$ and $\mathbb{P}(\phi_{lo}) = 1$, for all v, l, o. In the boundary models that posit no structural properties, we set $|\mathscr{T}| = 1$ and fix $\mathbb{P}(\pi_{1f}) = 1$, $\mathbb{P}(\lambda_{v1}) = 1$, and $\mathbb{P}(\omega_{to}) = 1$. Abusing notation, we refer to these models by saying $|\mathscr{L}| = 0$ and $|\mathscr{T}| = 0$, respectively.

Importantly for interpreting our results, having some lexical or structural type does not preclude having another: predicates and/or structures may have multiple lexical or structural types, respectively, or none at all. Thus, in this setup, sentences that disallow either future- or past-orientation can be explained in qualitatively different ways. In a model that admits lexical types, a particular predicate may fail to have a lexical type that licenses either orientation—including by having no lexical types at all. And in a model that admits structural types, a predicate may fail to select a type that licenses either orientation—including selecting no types at all. The upshot is that the interpretation of structural types in our model is importantly distinct from that in White and Rawlins' model (but analogous to An and White's), since not selecting any structural types does not mean not having any type signature. Rather, it is important to interpret the *type configurations*—i.e. the range of types a predicate or structure is and is not associated with—as defining *lexical* or *structural classes*, of which there are $2^{|\mathscr{L}|}$ and $2^{|\mathscr{T}|}$, respectively.

4.3 Model fitting

Analogous to White and Rawlins, we use the MegaOrientation dataset to obtain a measure of $\mathbf{A}$, modeling acceptability judgments for verb v in frame f with orientation o using an ordinal mixed model with unconstrained random cutpoints for each participant and $\mathbb{P}(a_{vfo})$ as a fixed effect. This allows us to estimate $\mathbb{P}(\lambda_{vt})$, $\mathbb{P}(\pi_{tf})$, $\mathbb{P}(\psi_{vl})$, $\mathbb{P}(\phi_{lo})$, and $\mathbb{P}(\omega_{to})$ as $\mathbb{P}(a_{vfo}) = 1 - \prod_{t,l} 1 - \mathbb{P}(\lambda_{vt})\mathbb{P}(\pi_{tf})\mathbb{P}(\psi_{vl})\mathbb{P}(\phi_{lo})\mathbb{P}(\omega_{to})$ (see An and White 2020).

To adjust for the fact that our method for selecting predicate-complement pairs for inclusion in MegaOrientation retains some variability in the underlying acceptability of each pair (see Footnote 2), we weight the log-likelihood of our model by the acceptability from MegaAttitude. Specifically, we normalize that acceptability using the ordinal model-based normalization described in White and Rawlins in press, standardize those values, and squash those standardized values to (0, 1) with the normal cumulative density function. We then use the resulting quantity for each predicate-complement pair to weight the likelihood for any item in MegaOrientation containing that pair.

Our model comparison procedure follows An and White closely. As such, the following description is a lightly modified version of theirs. Our main goal is to find the optimal settings, relative to our temporal orientation data, for (i) the number $|\mathscr{L}|$ of lexical properties relevant to temporal orientation; and (ii) the number $|\mathscr{T}|$ of structural properties relevant to temporal orientation. Higher values for $|\mathscr{L}|$ (with a fixed $|\mathscr{T}|$) or $|\mathscr{T}|$ (with a fixed $|\mathscr{L}|$) will necessarily fit the data as well or better than lower values, since a model with larger $|\mathscr{L}|$ or $|\mathscr{T}|$ can subsume the model with a smaller value. However, this better fit comes at the cost of increased risk of overfitting due to the inclusion of superfluous dimensions.

To mitigate the effects of overfitting, we conduct a five-fold cross-validation and select the model(s) with the best performance (in terms of our weighted loss) on held-out data. In this cross-validation, we pseudorandomly partition sentences from the temporal orientation experiment into five sets (folds), fit the model with some setting of $|\mathscr{L}|, |\mathscr{T}|$ to the acceptability responses for sentences in four of these sets (80% of the data), then compute the loss on the held-out set—repeating with each partition acting as the held-out set once. The assignment of items to folds is pseudorandom in that each fold is constrained to contain at least one instance of a particular verb with a particular complement type with some orientation. If such a constraint were not enforced, on some folds, the model would have

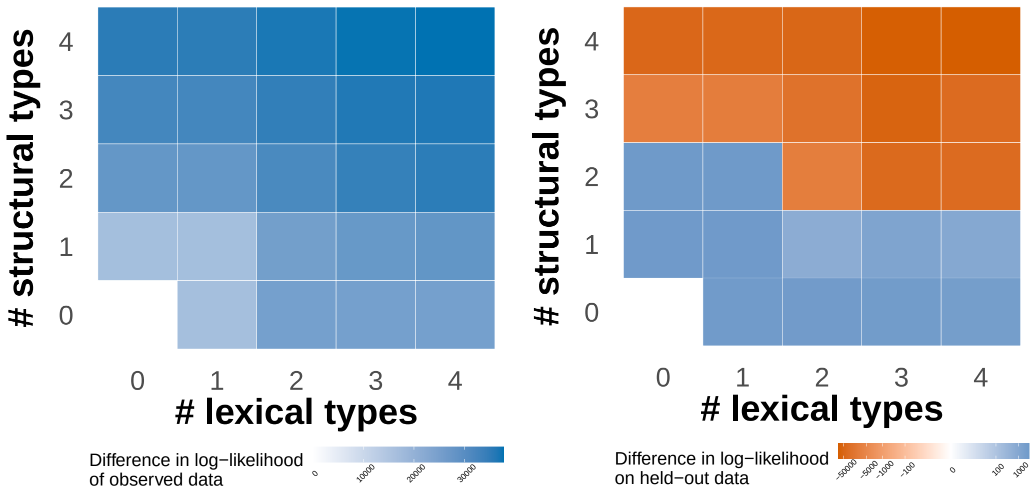

Figure 2: Comparison of model performance on observed data (left) and held-out data (right) against performance of baseline model ($|\mathscr{L}| = |\mathscr{T}| = 0$).

no data upon which to predict that verb with that complement. We consider each possible pairing of $|\mathscr{L}|, |\mathscr{T}| \in \{0,1,2,3,4\}$. The same partitioning is used for every setting of $|\mathscr{L}|$ and $|\mathscr{T}|$, enabling paired comparison by sentence.

We use the $|\mathscr{L}| = |\mathscr{T}| = 0$ model—i.e. the model positing that neither lexical nor structural factors contribute to temporal orientation—as the baseline model against whose performance the other models' performance is compared. By construction, this model always predicts the mean response across all predicates and frames for a particular orientation, and so models with $|\mathscr{L}| > 0$ or $|\mathscr{T}| > 0$ will always perform as well or better on data they have been exposed to, though not necessarily on data they have not been.

4.4 Results

Figure 2 plots the performance of each model we fit in terms of the difference between the log-likelihood it assigns to some data and the log-likelihood the baseline model ($|\mathscr{L}| = |\mathscr{T}| = 0$) assigns to those data. The left plot shows this comparison on the data that the model was fit to; the right plot shows this comparison on data it has not seen (held-out).

As expected, for a fixed setting of $|\mathscr{L}|$, the performance on observed data improves as $|\mathscr{T}|$ goes up; and for a fixed setting of $|\mathscr{T}|$, the performance on observed data improves as $|\mathscr{L}|$ goes up. More interestingly, we observe a clean division between models which perform better than baseline (blue) in the right plot and models which perform worse (orange). Of the better-performing models, two are structural ($|\mathscr{L}| = 0, 0 < |\mathscr{T}| < 3$), four are lexical ($|\mathscr{T}| = 0, |\mathscr{L}| > 0$), and five are mixed ($|\mathscr{T}| = 1$ or $|\mathscr{L}| = 1, |\mathscr{T}| = 2$).

The numerically best-performing model is the structural model with $|\mathscr{T}| = 2$ (ΔLL=2807, 95% CI=[2535, 3802]).[3] This model performs reliably better than all lexical models including the numerically best-performing one with $|\mathscr{L}| = 1$ (ΔLL=2277, 95% CI=[2510,

[3]Confidence intervals are computed using a nonparametric bootstrap ($n = 999$) of the mean difference in held-out likelihood across folds.

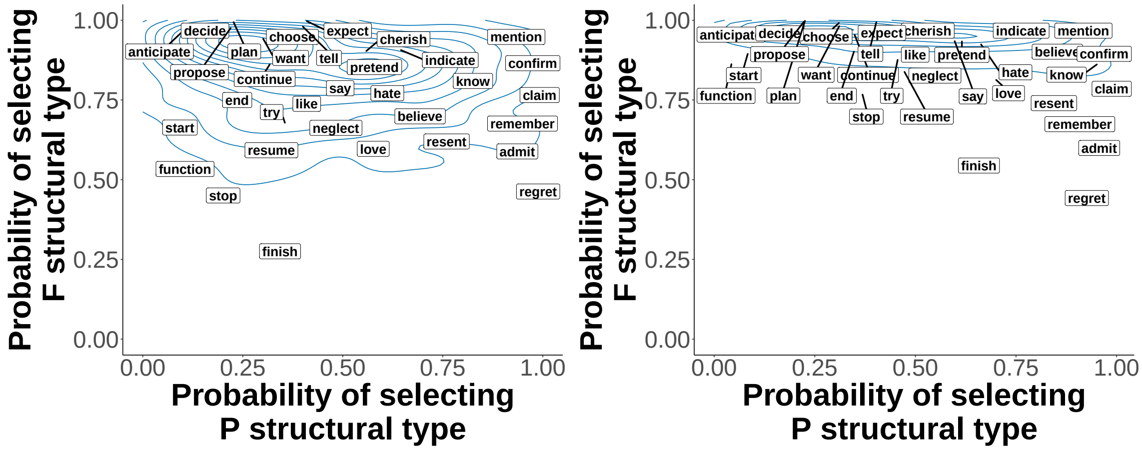

Figure 3: The probability that predicates selects P structural type (x-axis) or F structural type (y-axis) in the structural model (left) and the mixed model (right).

2759]), but it does not perform reliably better on held-out data than the best-performing mixed model with $|\mathscr{L}| = 1, |\mathscr{T}| = 2$ (ΔLL=2798, 95% CI=[2532, 3054]). Thus, as a matter of parsimoniously explaining the temporal orientation data at hand, a structural or mixed model should be preferred to a lexical model. We analyze both model types below.

5. Analysis

The results reported above suggest that positing more than two structural types is not warranted by the data. This finding might appear to speak against accounts that posit more structural types (Wurmbrand 2014), but caution is needed here for a couple reasons. First, because some structural types may be indistinguishable on the basis of their orientation alone, we may just be missing a distinction that might otherwise show up with a model that incorporates more information about the syntactic distribution in the types. Second, as discussed at the end of §4.2, $|\mathscr{T}| = 2$ implies $2^{|\mathscr{T}|}$ structural classes, and so it is important to consider how many such classes are actually selected across predicates.

In investigating the best-performing structural model ($|\mathscr{T}| = 2$) and mixed model ($|\mathscr{T}| = 2, |\mathscr{L}| = 1$), we find that they both associate one structural property with past orientation near probability 1 and future orientation near probability 0 and the other with future orientation near probability 1 and past orientation near probability 0. We refer to the former as the *P (structural) type* and the latter as the *F (structural) type*. All structures are associated with the F type near probability 1, but vary in their association with the P type: [NP __ VPing] is associated with the P type near probability 1, but [NP __ to VP[+EV]] and [NP was __ed to VP[+EV]] are associated with it at probability ∼0.4, with [NP __ to VP[-EV]] and [NP was __ed to VP[-EV]] in between the two. This pattern is consistent with the one observed in Figure 1, wherein [NP __ VPing] clearly allows past orientation and [NP __ to VP[+EV]] and [NP was __ed to VP[+EV]] nearly universally disallow it.

Figure 3 plots the probability that predicates select the F and P structural types. We see that the vast majority of predicates select the F type with high probability, while a (large) minority select the P type. Those that only select the F type with high probability—e.g. *de-*

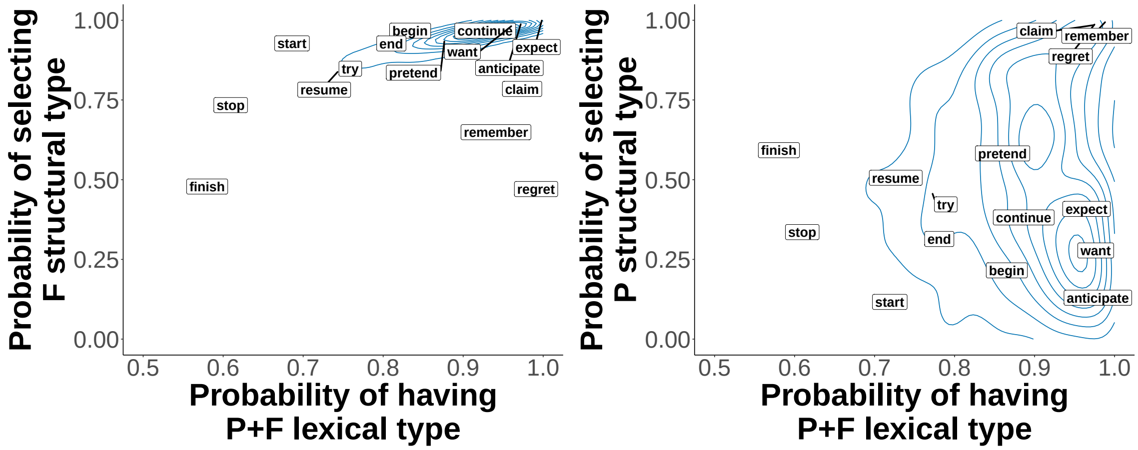

Figure 4: The probability that predicates have the P+F lexical type (*x*-axis) and select (*y*-axis) either the F structural type (left) or the P structural type (right).

cide, *plan*, *expect*, *want*—are those that are generally taken to either only give rise to future orientation across the nonfinite complements they select or that are at least compatible with such an orientation with each such complement. Those that select the P type but not the F type with high probability—e.g. *remember*, *regret*, *resent*—tend to allow past orientation in the gerundive complement but may only marginally allow future orientation with at least one of the infintival complements. Finally, those that select both with high probability tend to be those that allow both orientations with the gerundive complement—e.g. *confirm* and *mention*—though there are some surprising outliers—e.g. *claim* and *pretend*—which are generally implicated in sentences containing them being present-oriented.

These outliers contrast with other predicates commonly implicated in present orientation, like aspectuals (e.g. *start* and *stop*). The treatment of these predicates is the one place where the structural and mixed models appear to pull apart in terms of selection. In the structural model, these predicates select neither type with high probability, implying that they only occur in present-oriented sentences. It is exactly the distinction between propositional attitude predicates and aspectuals that accounts like Wurmbrand's attempt to capture, and so this behavior might indicate that the model has picked up on the relevant distinction.

In contrast, in the mixed model, these predicates tend to select the the F type with relatively higher probability, thereby blurring the selectional distinction observed in the structural model. Instead, the mixed model appears to capture this distinction in which predicates take the single lexical type posited by the model. This lexical type is associated with high probability ($\sim$1) for being compatible with both past and future orientation, and thus we refer to it as the *P+F lexical type*. Figure 4 plots the probability that different predicates have the P+F lexical type (*x*-axis) and select either structural type (*y*-axis).

We see two interesting patterns. First, aspectuals tend to show lower probability of having the P+F lexical type, implying that these predicates only occur in present-oriented sentences. Thus, in contrast to the structural model, the mixed model attributes the fact that aspectuals tend to show up in present-oriented sentences to the predicate itself, rather than its complement, though it associates past and future orientation with the complement. Second, we see that the probability of having the P+F lexical type is highly correlated with the probability of having the F structural type (Spearman's ρ=0.76, p<0.001) but not the P structural type (Spearman's ρ=-0.02, p=0.49). This correlation might suggest some connection between whatever aspect of the predicate denotation the P+F lexical type is tracking and whatever aspect of the complement denotation the F structural type is tracking.

6. Conclusion

We approached the question of what gives rise to the temporal orientation of sentences containing nonfinite embedded clauses—the clause-embedding predicate, its nonfinite complement clauses, or both—using a combination of lexicon-scale data collection and quantitative model comparison. We found that, on the basis of temporal orientation data alone, models that attribute temporal orientation only to predicate denotations reliably underperform those that attribute it to only complement denotations or both predicate and complement denotations on unseen data. This finding suggests that a purely lexical approach to temporal orientation should be dispreferred relative to a structural or mixed approach.

But caution is warranted here. The best-performing lexical model only slightly underperforms the best-performing structural model (relative to the baseline), and so it may turn out that, if we were to incorporate additional aspects of interpretation, such as construal (Landau 2000, White and Grano 2014, Grano 2015, Pearson 2016), into our model the lexical models may outperform the structural and mixed models. Given additional lexicon-scale data relevant to said aspect of interpretation, the modeling paradigm we develop in this paper can be straightforwardly extended for exploring this possibility. One way to do this is to keep constant the components of the model that deal with (i) which structural types a predicate selects; (ii) which lexical types it has; and (iii) which structural types a structure can realize, but to allow for multiple interpretive components, each mapping from the lexical and/or structural types to the lexicon-scale data relevant to that phenomenon. We believe this sort of extension of our model is a potentially fruitful direction for future work.

References

Abusch, Dorit. 2004. On the Temporal Composition of Infinitives. In *The Syntax of Time*, ed. Jacqueline Gueron and Jacqueline Lecarme, 27–53. MIT Press.

An, Hannah, and Aaron White. 2020. The lexical and grammatical sources of neg-raising inferences. *Proceedings of the Society for Computation in Linguistics* 3:220–233.

Bhatt, Rajesh. 1999. Covert Modality in Non-Finite Contexts. Doctoral dissertation, University of Pennsylvania.

Bresnan, Joan W. 1972. Theory of Complementation in English Syntax. Doctoral dissertation, Massachusetts Institute of Technology.

Grano, Thomas Angelo. 2015. *Control and restructuring*. Oxford: Oxford University Press. OCLC: 979110141.

Grano, Thomas Angelo. 2017. Control, temporal orientation, and the cross-linguistic grammar of trying. *Glossa: a journal of general linguistics* 2:94.

Grimshaw, Jane. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10:279–

326.

Karttunen, Lauri. 1971. Implicative verbs. *Language* 340–358.

Katz, Graham. 2001. (A)temporal complements. In *Audiator Vox Sapientiae*, ed. C. Fery and W. Stemefeld, 240–258. Berlin: Akademie Verlag.

Landau, Idan. 2000. *Elements of Control: Structure and Meaning in Infinitival Constructions*. Studies in Natural Language and Linguistic Theory. Dordrecht: Kluwer Academic Publishers.

Martin, Roger. 2001. Null Case and the Distribution of PRO. *Linguistic Inquiry* 32:141–166. Publisher: The MIT Press.

Moulton, Keir. 2009. Natural Selection and the Syntax of Clausal Complementation. Doctoral dissertation, University of Massachusetts, Amherst.

Ogihara, Toshiyuki. 1989. Temporal Reference in English and Japanese. PhD Thesis, University of Texas at Austin.

Ogihara, Toshiyuki. 1996. *Tense, Attitudes, and Scope*. Studies in Linguistics and Philosophy. Springer Netherlands.

Pearson, Hazel. 2016. The Semantics of Partial Control. *Natural Language & Linguistic Theory* 34:691–738.

Pesetsky, David. 1982. Paths and Categories. Doctoral dissertation, Massachusetts Institute of Technology.

Pesetsky, David. 1991. Zero syntax: vol. 2: Infinitives.

Pesetsky, David, and Esther Torrego. 2004. Tense, case, and the nature of syntactic categories. In *The Syntax of Time*, ed. Jacqueline Gueron and Jacqueline Lecarme, 27–53. Cambridge, MA: MIT Press.

Stowell, Tim. 1982. The tense of infinitives. *Linguistic Inquiry* 13:561–570.

White, Aaron Steven. 2014. Factive-implicatives and modalized complements. In *Proceedings of the 44th annual meeting of the North East Linguistic Society*, ed. Jyoti Iyer and Leland Kusmer, 267–278. University of Connecticut.

White, Aaron Steven, and Thomas Grano. 2014. An experimental investigation of partial control. In *Proceedings of Sinn und Bedeutung*, ed. U. Etxeberria, A. Falaus, A. Irurtzun, and B. Leferman, volume 18, 469–486.

White, Aaron Steven, and Kyle Rawlins. 2016. A computational model of S-selection. *Semantics and Linguistic Theory* 26:641–663.

White, Aaron Steven, and Kyle Rawlins. 2018. The role of veridicality and factivity in clause selection. In *Proceedings of the 48th Annual Meeting of the North East Linguistic Society*, to appear. Amherst, MA: GLSA Publications.

White, Aaron Steven, and Kyle Rawlins. in press. Frequency, Acceptability, and Selection: A case study of clause-embedding. *Glossa* .

Williamson, Gregor. 2017. Doxastic Diversity and Covert Temporal Operators. *UCLWPL* 29:42–58.

Williamson, Gregor. 2019. The temporal orientation of infinitives. In *Proceedings of Sinn und Bedeutung*, volume 23, 461–478. Issue: 2.

Wurmbrand, Susi. 2001. *Infinitives: Restructuring and Clause Structure*. De Gruyter.

Wurmbrand, Susi. 2014. Tense and aspect in English infinitives. *Linguistic Inquiry* 45:403–447.