

# UC Merced

## Proceedings of the Annual Meeting of the Cognitive Science Society

### Title

A Factorized Probabilistic Model of the Semantics of Vague Temporal Adverbials Relative to Different Events

### Permalink

<https://escholarship.org/uc/item/623599f4>

### Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

### Authors

Kenneweg, Svenja

Deigmoeller, Joerg

Eggert, Julian

et al.

### Publication Date

2025

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# A Factorized Probabilistic Model of the Semantics of Vague Temporal Adverbials Relative to Different Events

Svenja Kenneweg (skenneweg@techfak.uni-bielefeld.de)

Bielefeld University, Faculty of Technology, Universitätsstraße 25, 33615 Bielefeld, Germany

Jörg Deigmoeller (Joerg.Deigmoeller@honda-ri.de)

Honda Research Institute Europe, Carl-Legien-Straße 30, 63073 Offenbach am Main, Germany

Julian Eggert (Julian.Eggert@honda-ri.de)

Honda Research Institute Europe, Carl-Legien-Straße 30, 63073 Offenbach am Main, Germany

Philipp Cimiano (cimiano@techfak.uni-bielefeld.de)

Bielefeld University, Faculty of Technology, Universitätsstraße 25, 33615 Bielefeld, Germany

## Abstract

Vague temporal adverbials, such as “recently,” “just” and “long time ago,” describe the temporal distance between a past event and the utterance time, but leave the exact duration underspecified. In this paper, we introduce a factorized model that captures the semantics of these adverbials as probabilistic distributions. These distributions are composed with event-specific distributions to yield a contextualized meaning for an adverbial applied to a specific event. We fit the model’s parameters using existing data capturing judgements of native speakers regarding the applicability of these vague temporal adverbials to events that took place a given time ago. Comparing our approach to a non-factorized model based on a single Gaussian distribution for each pair of event and temporal adverbial, we find out that, while both models have similar predictive power, our model is preferable in terms of Occam’s razor, as it is simpler and has a better extendability.

**Keywords:** compositional modeling; temporal adverbials; time cognition; probabilistic distributions; semantic vagueness

## Introduction

Vague temporal adverbials such as *recently*, *just*, *some time ago*, or *long time ago* are used to describe the temporal recency or distance between an event that took place in the past and the utterance time. These temporal adverbials are vague in the sense that they do not specify the exact length of the temporal interval passed between the event in question and the utterance time, leaving this length under-specified. Yet, this length is not arbitrary, as corroborated by the following examples.

While we can say ‘*I just celebrated my birthday*’ when it was a few days ago, we would rather not utter such a sentence when our birthday was months or years ago. Take another example ‘*I just brushed my teeth*’. Such a sentence can be uttered when the event of brushing our teeth is minutes, but not days or weeks, ago.

Although both examples use the same vague temporal adverbial *just*, the length of the temporal interval between the events and the utterance time varies, ranging from a few minutes to several days. This suggests that this length is shaped not only by the choice of the temporal adverbial, but also by the event itself. In fact, it has been argued that there are two crucial variables of an event that affect this length: duration and frequency (Van Jaarsveld & Schreuder, 1985).

These insights indicate that a successful approach to modeling the semantics of vague temporal adverbials must be compositional, capturing not only how these adverbials define a temporal frame of reference but also how different events shape the resulting interpretation.

In this paper we present such a compositional approach to modeling the semantics of vague temporal adverbials. Grounded in the theory that the semantics of vague expressions can be viewed as having crisp yet uncertain boundaries (Lawry & Tang, 2009; Williamson, 1998), we represent the vague temporal adverbials and their corresponding events as probabilistic distributions. This allows us to assign higher probabilities to those timepoints that are most likely to be described by a given vague temporal adverbial and corresponding event. As it would not be realistic to assume that each vague temporal adverbial, such as *just*, has one distinct meaning for each event, we factorize the interpretation: the meaning of an adverbial distribution remains fixed, and the event properties shape how that distribution is applied.

Therefore our compositional model relies on two types of probability density functions  $P_{Adv}$  and  $P_{Ev}$ . When applied to an event that has happened  $t$  times units ago in the past,  $P_{Ev}(t)$  transforms the distance between the `reference_time` –  $t$  into a distribution specific for the event  $Ev$  that represents the probability that this event  $Ev$  that took place  $t$  time units ago can actually be said to have happened *before* the `reference_time`. The `reference_time` is the temporal anchor point relative to which the occurrence of the event is evaluated. The probability density function  $P_{Adv}$  takes the result of the transformation by  $P_{Ev}$  so that the distribution  $P_{Adv}(P_{Ev}(t))$  then models the likelihood that the vague temporal adverbial  $Adv$  can be felicitously used to describe an event  $Ev$  that happened  $t$  time units ago.

The main contribution of our paper is the factorized compositional model to capturing the meaning of vague temporal adverbials, which is described in the Section “Compositional Model”. We fit the parameters of the model relying on data from Kenneweg, Jackson, Deigmoeller, Eggert, and Cimiano (2024), in which we elicited the judgments of English speakers regarding the applicability of the different vague temporal

adverbials to describing events that are a particular time unit away from the utterance time. We compare our approach to a non-factorized approach as baseline that models the likelihood for every pair of adverb and event explicitly in terms of data fit (accuracy) and extendability.

## Background

Vague predicates are characterized by the fact that there are borderline cases, that is cases where it is not clear whether the vague predicate applies or not. For example, the adjective *tall* can be said to be vague as there are cases of people that can neither be clearly identified as *tall* nor *not tall*. Philosophically, vagueness has been characterized by reference to Sorites Paradox, which (instantiated for the case of tallness) would claim that anyone who is merely 1mm shorter than a tall person intuitively counts as tall. Yet, if we start with a 2.5m person who is clearly tall, this principle leads us, by a series of small steps, to the absurd conclusion that a person who is only 10mm high would also count as tall.

While vagueness occurs across parts-of-speech, much of the literature has focused on adjectives (H. Kamp & Sassoon, 2016; Kennedy, 2007; Lassiter & Goodman, 2014; Solt & Gotzner, 2012). H. Kamp and Sassoon (2016) have proposed a categorization of vague adjectives into *relative* (e.g., tall, short) and *absolute* adjectives (e.g., clean, dirty, straight). Beyond adjectives, other parts-of-speech, like quantifiers (*many*, *few*) and nouns (*mountain*) also exhibit vagueness. In this paper, we consider the case of vague temporal adverbials such as *recently*, *just*, *some time ago*, and *long time ago*, which so far have not received substantial attention in the literature on vagueness. While we focus on this particular case, our compositional approach could nevertheless be generalized to model vagueness compositionally in general, for instance, to cases of spatial vagueness.

There are two prominent, fundamentally different philosophical frameworks of how to treat vagueness. The *Epistemic theory* (Williamson, 1998) assumes that vague predicates have precise but unknown boundaries, and that vagueness arises due to a lack of knowledge rather than inherent indeterminacy in the world itself. It posits that a predicate either applies or does not apply to a specific case, following a binary distinction with a certain probability. This probability depends on the degree of epistemic certainty about the threshold, which can be modeled using probability theory. In this context, the epistemic probability quantifies, e.g., the uncertainty that a specific individual's height falls above or below the unknown threshold for being classified as "tall." As the available information increases, the epistemic probability can shift, reflecting greater confidence in determining whether a person qualifies as tall or not (for a more formal account, see Climenhaga (2020)). Therefore, the epistemic probability captures the agent's belief about the truth of the predicate, given the evidence at hand. This approach emphasizes that the vagueness is not an inherent feature of the concept itself,

but rather a result of limited knowledge about the exact cut-off. Following the idea that predicates have crisp boundaries, *Supervaluationism theory*, as proposed by Fine (1975), uses a logical approach and models subjectivity in vague predicates by assuming that their extension varies across models. A person may be tall in some models of a theory, but not-tall in others. This leads to the definition of statements that are *supertrue* (true in all models) and *superfalse* (false in all models). While the *Supervaluationism theory* captures logical subjectivity, it does not address where boundaries lie or how speakers determine predicate applicability in specific contexts.

In contrast to the approaches that assume that boundaries are crisp but unknown, there are approaches that attempt to capture the degree to which a vague predicate applies to a certain case. The most prominent approach along these lines, *Fuzzy Set Theory* (Zadeh, 1965), allows for different degrees of membership of an instance in a *fuzzy set* along the closed interval  $[0, 1]$ . A taller person might thus be member of the fuzzy set of '*tall persons*' to a degree of 0.8.

Along these lines, *Prototype Theory* is a further approach that models the degree of membership with respect to some prototypical instance of the vague predicate (H. Kamp & Partee, 1995). According to this theory, some instances of a concept are more representative than others; the most representative instances are the prototypes. The membership degree of an instance can then be defined by a similarity function to the prototype of the corresponding concept (Decock & Douven, 2014). Consequently, the probability of an object belonging to a concept decreases with the distance to the prototype of this concept. This relationship can be modeled with a Gaussian distribution (Bouraoui & Schockaert, 2018). The *conceptual spaces* theory by Gärdenfors (2000) also allows for the modeling of vague concepts in terms of convex geometrical regions in a metric space that allows to define the closeness of instances to the concept (Doven, 2016).

Independently of which approach is followed, an important property of vague predicates is that their meaning is context-dependent. For instance, a baby is considered *young* at the age of a few months but not at a few years, whereas an adult could still be considered *young* at the age of 20 (Damerau, 1977). According to J. A. W. Kamp and Keenan (1975), particularly gradable adjectives like *big* or *small* should be interpreted relative to a reference class.

Prior work has emphasized the necessity of a reference or comparison class in the interpretation of vague predicates, particularly adjectival ones. Heim (2000) discusses how adjectives access a reference class through their degree argument (DegP). For instance, in the sentence "Every girl is taller than 4 feet," the phrase "than 4 feet" provides the comparison standard against which the adjective taller is evaluated. Schmidt, Goodman, Barner, Edu, and Tenenbaum (2010) propose two distinct types of models for the interpretation of *tall*: a threshold-based model, where an object is considered tall if it exceeds a contextually defined threshold, and a category-

based model, where objects are assigned to clusters based on shared statistical properties such as height. Similarly to the threshold-based model, Kennedy (2007) posits a context-dependent "standard of comparison" which relies on a reference class. In "big dog" the reference class is "dog" and in "big children" "children". This reference class sets the threshold above which a dog or children is considered big and can be determined by using typical or average values of the reference class. Qing and Franke (2014) and Lassiter and Goodman (2017) have introduced probabilistic models that capture how speakers and listeners pragmatically reason about vague predicates in relation to comparison classes.

While these approaches offer valuable insights into the context-sensitivity of vague adjectives, they do not adopt a fully compositional perspective in the sense of Frege's principle of compositionality (Frege, 1953), where complex meanings are systematically built from the meanings of their parts and their mode of combination. Moreover, in these accounts, the reference class is typically treated as a statistical distribution over a population of instances, rather than as something structured in its interaction with a specific event or object.

In the context of vague temporal adverbials such as we consider in this paper, the reference class relates to the different events that they can apply to. Previous work has identified that key parameters of events determining in how far they can be described by a certain vague temporal adverbial are the characteristic *frequency* and *duration* of the event (Van Jaarsveld & Schreuder, 1985). This suggests that events that have a similar frequency and duration would provide a similar or comparable reference class.

Building on this insight, we model the reference class by modeling the *characteristic temporal signature* of events and making it comparable so that events with a similar frequency and duration are closer in the temporal signature space. The meaning of a vague temporal adverbial applied to a specific event results then from the compositional application of its specific signature to the *characteristic temporal signature* of the event.

Thus, our work extends previous probabilistic accounts of vagueness by introducing a compositional semantic structure that explicitly models how the meaning of vague expressions arises from the structured interaction between events and adverbials. This offers a novel framework for modeling vagueness not merely at the level of context sensitivity, but at the level of compositional meaning construction.

## Compositional Model

The model we introduce is a factorized model, following Frege's principle of compositionality (Frege, 1953), paraphrased as "*The meaning of a complex predicate can be modeled via the meaning of its parts and how they are composed together*". In our framework, these parts consist of the vague temporal adverbial (e.g. *just, recently, ...*) on the one hand and their reference class - the event (e.g. *Brushing Teeth, Sabbatical, ...*) on the other.

In our case of modeling the complex compositional meaning of a vague temporal adverbial and event, there are at least three important 'parts' to consider:

- Each temporal adverbial (e.g. *just, recently, long time ago*) conveys a distinct notion of temporal proximity - thus having a distinct meaning
- Events can be classified based on their *characteristic temporal signature*, which is determined by factors such as duration and frequency (Van Jaarsveld & Schreuder, 1985)
- The number  $t$  of time units that have passed since the specific event happened until the time of utterance (`reference_time`)

As we follow the *epistemic theory* proposed by Williamson (1998), we take the application of a particular vague temporal adverbial  $Adv$  to an event  $Ev$  that was  $t$  time units ago to have crisp but unknown boundaries. We model the uncertainty in the boundaries using probability theory, and propose a factorized probabilistic model that relies on two probability density functions:  $P_{Adv}$  and  $P_{Ev}$ .  $P_{Adv}$  is specific for each adverbial  $Adv$  and  $P_{Ev}$  for each event  $Ev$ .

The factorized model, which returns the probability that the adverbial  $Adv$  applies to an event  $Ev$  that happened  $t$  time units ago, is as follows:

$$P_{Adv}(P_{Ev}(t))$$

Intuitively, the (probability) function  $P_{Ev}$  models the probability that an event  $Ev$  that happened  $t$  time units *before* the `reference_time` can actually be said to have happened *before* the `reference_time`. Take the example of a teeth brushing event: a teeth brushing event that happened a few minutes ago has, from a probabilistic point of view, clearly happened before now (as now is here the `reference_time`), while a teeth brushing event that happened a few milliseconds ago did less clearly happen *before* now. The probabilistic distribution  $P_{Ev}$  thus captures the temporal signature of events  $Ev$  with respect to the likelihood that we are willing to describe instances of the event as having happened *before* depending on how long ago they actually happened. This event specific probability distribution models the reference class, as different events will generally have a different signature of this function. The function models the uncertainty in describing that one event  $Ev$  occurs before some other time point or interval.

Assuming that there is some uncertainty for each event in terms of when it took place exactly, we model events as Gaussian distributions of the following form:

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right)$$

Conversely, the uncertainty related to whether an event  $Ev$  took place before another event or time interval can also be modeled as a Gaussian distribution as follows, where *erf* is the Gauss error function:

$$\frac{1}{2} \left( \operatorname{erf} \left( \frac{\mu}{\sqrt{2}\sigma} \right) + 1 \right)$$

We thus model the function  $P_{Ev}$  as a Gaussian distribution of the above type, where  $\mu$  represents the number  $t$  of time units that have passed since the specific event occurred. The parameter  $\sigma$  varies depending on the event, influencing the function to output higher probabilities for time units that are closer to or farther away from the utterance time. The second function of our model, the adverbial-specific function,  $P_{Adv}$ , takes this normalized value as input. As mentioned before, we follow Bouraoui and Schockaert (2018) and model each adverbial as a Gaussian distribution around its prototype. Thus, the adverbial-specific function  $P_{Adv}$  is expressed as follow:

$$\exp \left( -\frac{1}{2} \left( \frac{x - \mu_a}{\sigma_a} \right)^2 \right)$$

This function has the two parameters  $\mu_a$  and  $\sigma_a$ , which are the mean and standard deviation of the Gaussian distribution and defined individually for each adverbial.

Thus, the *event-specific function*  $P_{Ev}$  feeds into a normalized Gaussian distribution with specific parameters for each temporal adverbial. This final step yields a probability representing how likely it is that a human speaker would use a certain temporal adverbial, based on the probability of the described event’s temporal precedence.

### Model Fitting

We fit the parameters of the model using data from our previous study that elicited judgements of native speakers regarding the applicability of a vague temporal adverbial to an event that took place  $t$  time units ago (Kenneweg et al., 2024). Through an online survey, we collected probabilities reflecting the likelihood of a specific temporal adverbial for a given event and elapsed time since the event occurred. For this purpose, our participants, which were all adult native English speakers, rated the applicability of a temporal adverbial applied in english sentences to a past event on a Likert Scale. An example sentence from the survey is: “Tom’s Birthday was 1 day ago. Statement: Tom had his birthday recently.”

The study comprised 6 events (*Brushing Teeth*, *Birthday*, *Vacation*, *Sabbatical*, *Year Abroad*, *Marriage*) and 4 vague temporal adverbials (*just*, *recently*, *some time ago*, *long time ago*). For each event - adverbial pair, we explored its applicability across 7 elapsed times at which the event occurred. We divided the survey into 6 separate questionnaires, each targeting one event applied to all adverbials. Each questionnaire was completed by 100 participants recruited via the Prolific platform<sup>1</sup>.

In order to fit the parameters of our model, we normalized their Likert-scale responses into the interval  $[0, 1]$ . This

results in 16.800 data points (6 events x 4 adverbials x 7 times x 100 votes). Given the compositional nature of our model, we solve the combined function  $P_{Adv}(P_{Ev}(t))$  for each event-adverbial pair, but over all data points - therefore for example finding the best parameters of *Just* over all evaluated events. For instance, for the event *Vacation*, this involved solving equations like  $P_{Just}(P_{Vacation}(1Month)) = 0.2$ . Time values were standardized by converting them to minutes, with *1Month* corresponding to 43800 minutes (based on an average month length of 30,4 days). For fitting, we used least-squares optimization to minimize the residuals, i.e., the difference between the observed and modeled probabilities.

From the fitting we get the optimal parameters (means and standard deviations) for each adverbial-specific function,  $P_{Adv}$ , as well as the standard deviation for each event-specific function  $P_{Ev}$  for all 4 fitted adverbials and 6 events. The final standard deviations for each event-specific function  $P_{Ev}$  are shown in Table 1, while the means and standard deviations for the Gaussian distributions of each adverbial-specific function  $P_{Adv}$  are presented in Table 2. To illustrate how these functions combine, Figure 1 shows the composition of all 6 event-specific functions with the 4 adverbial-specific functions.

Table 1: Standard deviation of each event-specific function in minutes

| Event          | $\sigma_e$ |
|----------------|------------|
| Brushing Teeth | 935        |
| Birthday       | 314830     |
| Vacation       | 396579     |
| Sabbatical     | 798494     |
| Year Abroad    | 1240803    |
| Marriage       | 2334869    |

Table 2: Standard deviation and mean of each adverbial-specific function

| Adverbial     | $\sigma_a$ | $\mu_a$ |
|---------------|------------|---------|
| Just          | 0.04       | 0.48    |
| Recently      | 0.09       | 0.45    |
| Some Time Ago | 0.19       | 0.78    |
| Long Time Ago | 0.23       | 1.00    |

### Evaluation

We evaluate our compositional model  $P_{Adv}(P_{Ev}(t))$  by comparing it with a baseline, a non-factorized model that relies on a single Gaussian distribution for each pair of vague temporal adverbial and event. We compare the two models in terms of two key aspects: **accuracy** and **extendability**. In our case, the non-factorized model is a normalized Gaussian function of the following form:

<sup>1</sup><https://www.prolific.com/>

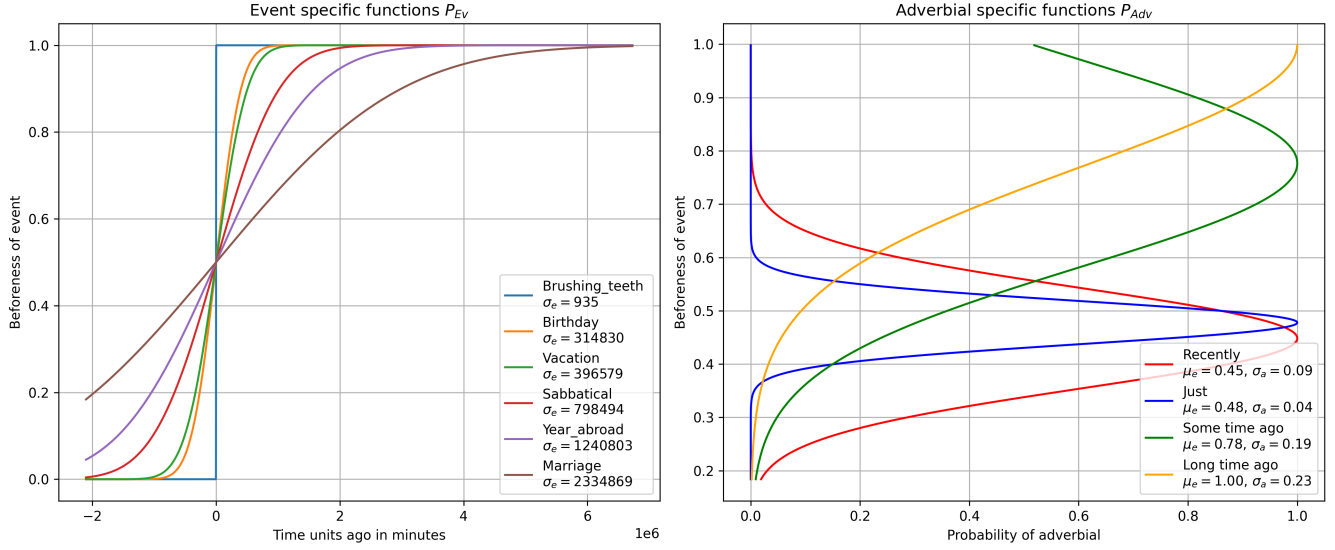


Figure 1: Event-specific functions and adverbial-specific functions for all fitted events and adverbials.

$$\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

The Gaussian for the non-factorized model is fitted onto the same empirical data from Kenneweg et al. (2024) via least squares optimization, but with one difference: each event-adverbial pair is fitted its own Gaussian, such as one for the event *Vacation* and the adverbial *recently*, another one for the same event with the adverbial *some time ago*, and so on.

### Accuracy

Accuracy measures how closely the model’s predictions align with the empirical data from Kenneweg et al. (2024), which serve as ground truth data. To evaluate the accuracy, we predict an adverbial’s probability for a given elapsed time point of an event using our model and compute the average absolute difference between the predicted value and all votes from the empirical data. Table 3 shows this average prediction error per event, adverbial and over all events and adverbials for our model and the normalized Gaussian.

### Extendability

Extendability describes the ability of a system to adapt to novel scenarios or inputs. In our case, we are interested in understanding the effort involved in extending the model to new adverbials and/or events. The dataset used to fit the parameters of the model comprises 6 events and 4 vague temporal adverbials. For our compositional model,  $6 + 4 = 10$  fitted functions were needed to include all events and vague temporal adverbials from the empirical dataset. The normalized Gaussian requires one function for each event-adverbial pair, resulting in  $6 * 4 = 24$  functions.

Table 3: Average error per elapsed time point, computed across events and adverbials for both our compositional model (Factorized Model) and normalized Gaussian (Non-factorized Model).

| Type      | Name           | Factorized Model | Non-factorized Model |
|-----------|----------------|------------------|----------------------|
| Event     | Brushing Teeth | 0.19             | 0.22                 |
|           | Birthday       | 0.19             | 0.19                 |
|           | Vacation       | 0.19             | 0.19                 |
|           | Sabbatical     | 0.21             | 0.22                 |
|           | Year Abroad    | 0.21             | 0.21                 |
|           | Marriage       | 0.19             | 0.18                 |
| Adverbial | Just           | 0.16             | 0.15                 |
|           | Recently       | 0.18             | 0.17                 |
|           | Some time ago  | 0.28             | 0.29                 |
|           | Long time ago  | 0.18             | 0.20                 |
| Overall   |                | 0.20             | 0.20                 |

Expanding the compositional model to include one additional event would require only one additional function, yielding  $7 + 4 = 11$  functions in total. The normalized Gaussian approach would require one additional function for each contained adverbial, yielding 4 additional functions and  $7 * 4 = 28$  functions in total. Table 4 illustrates this scalability effect by representing the number of required functions for up to 16 events and 16 adverbials.

### Discussion

In developing this model, we have formally defined the *reference class*—the context in which a vague predicate is interpreted—in line with Frege’s principle of compositionality, by

Table 4: Number of needed functions for x events and y adverbials for the compositional model (Factorized Model) and normalized Gaussian (Non-factorized Model).

| Number of |            | Needed Functions |                      |
|-----------|------------|------------------|----------------------|
| Events    | Adverbials | Factorized Model | Non-factorized Model |
| 2         | 2          | 4                | 4                    |
| 2         | 4          | 6                | 8                    |
| 2         | 8          | 10               | 16                   |
| 2         | 16         | 18               | 32                   |
| 4         | 16         | 20               | 64                   |
| 8         | 16         | 24               | 128                  |
| 16        | 16         | 32               | 256                  |

combining a vague temporal adverbial with a specific event. With this approach we extend previous probabilistic accounts of vagueness, which do not model the reference class compositionally in the Fregean sense (Damerau, 1977; Heim, 2000; J. A. W. Kamp & Keenan, 1975; Kennedy, 2007; Schmidt et al., 2010; Qing & Franke, 2014; Lassiter & Goodman, 2017).

Our initial analysis reveals that each examined event has its own distinct standard deviation (see Table 1). This suggests that every event possesses a *characteristic temporal signature*, most likely determined by its frequency and duration (Van Jaarsveld & Schreuder, 1985). For instance, *brushing teeth* exhibits a relatively narrow standard deviation ( $\sigma_e = 935$ ), whereas *sabbatical* shows a considerably larger standard deviation ( $\sigma_e = 798494$ ). These findings indicate that the *characteristic temporal signature* shapes the reference class function (event function  $P_{Ev}$ ); however, further investigation is needed to understand precisely how these *characteristic temporal signature* exert their influence. Although the empirical dataset includes only 6 events - and more evaluation is necessary - our preliminary results (which match the accuracy of the baseline model; see Table 3) suggest that our compositional model effectively captures the meaning of vague temporal adverbials and is a good starting point for further research.

Following an argument rooted in Occam’s Razor principle, our factorized model is clearly preferable to a non-factorized model as it has similar predictive power while representing a simpler explanation of the phenomenon.

As shown in the ‘Evaluation’ Section, its core strength lies in its extendability: rather than requiring a dedicated function for each event-adverbial pair, our compositional approach separates the modeling of events and adverbials into two distinct components. In doing so, the needed number of functions for the compositional model increases linearly as the number of events and adverbials grows. In contrast, for the baseline model (the normalized Gaussian), the number of functions would increase quadratically. Concretely (as shown in Table 4), when modeling 16 events and 16 vague temporal adverbials, the normalized Gaussian would need  $16 * 16 =$

256 functions, while our model requires  $16 + 16 = 32$ —only about 12.5% of the baseline approach. Despite this clear benefit in terms of extendability, the accuracy of our model is comparable to the accuracy of the baseline approach: both have an averaged error of 0.2 (see Table 3).

While we have argued that our model is easier to extend compared to a non-factorized model, an inherent limitation of the model is that we would need to acquire additional data for new adverbials and events as in the study of Kenneweg et al. (2024). An interesting direction for future work is whether the generalizability of the model can be improved by considering equivalence classes or categories of events that share a similar duration and frequency.

## Conclusion & Future Work

We introduced a factorized compositional model that formally captures the semantics of vague temporal adverbials. Our approach models the application of a vague temporal adverbial to an event by defining two probability density functions:  $P_{Adv}$  for the Adverbial and  $P_{Ev}$  for the event. The resulting composition  $P_{Adv}(P_{Ev}(t))$  computes the probability that an adverbial applies to an event that happened  $t$  time units ago. In defining  $P_{Ev}$ , the reference class for the vague temporal adverbials, we extend previous approaches by formally defining a reference class of a vague predicate in line with Frege’s principle of compositionality.

By fitting our compositional model to empirical survey data from Kenneweg et al. (2024), we instantiated the model for 6 events and 4 vague temporal adverbials.

Our evaluation demonstrates that our compositional approach matches the predictive accuracy of a non-factorized baseline model while being simpler (less parameters) and easier to extend to new events and adverbials. However, it currently addresses only a limited range of 6 events and 4 vague temporal adverbials - far fewer than those encountered in everyday life and language. Although extending our model to additional events and adverbials requires only defining one function for each additional event or adverbial, further research is needed to find out how to systematically induce the functions’ parameters, possibly without collecting new data through surveys.

Kenneweg et al. (2024).

## References

- Bouraoui, Z., & Schockaert, S. (2018, 07). Learning conceptual space representations of interrelated concepts..
- Climenhaga, N. (2020, 11). The structure of epistemic probabilities. *Philosophical Studies*, 177, 1–30. doi: 10.1007/s11098-019-01367-0
- Damerau, F. J. (1977). On “fuzzy” adjectives. *Linguistics*, 15(196), 57–64. doi: doi:10.1515/ling.1977.15.196.57
- Decock, L., & Douven, I. (2014, 12). What is graded membership? *Nous*, 48. doi: 10.1111/nous.12003
- Douven, I. (2016, 06). Vagueness, graded membership, and conceptual spaces. *Cognition*, 151, 80-95. doi: https://doi.org/10.1016/j.cognition.2016.03.007

- Fine, K. (1975). Vagueness, truth and logic. *Synthese*, 30(3-4), 265–300. doi: 10.1007/bf00485047
- Frege, G. (1953). *The foundations of arithmetic: A logico-mathematical enquiry into the concept of number* (J. L. Austin, Trans.). Oxford: Blackwell.
- Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. The MIT Press.
- Heim, I. (2000, 09). Degree operators and scope. *Proceedings from Semantics and Linguistic Theory*, 10. doi: 10.3765/salt.v10i0.2722
- Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition*, 57(2), 129-191. doi: [https://doi.org/10.1016/0010-0277\(94\)00659-9](https://doi.org/10.1016/0010-0277(94)00659-9)
- Kamp, H., & Sassoon, G. W. (2016). Vagueness. In M. Aloni & P. Dekker (Eds.), *The cambridge handbook of formal semantics*. Cambridge University Press.
- Kamp, J. A. W., & Keenan, E. L. (1975). Two theories about adjectives. In *Formal semantics of natural language*. Cambridge University Press.
- Kennedy, C. (2007, 02). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30. doi: 10.1007/s10988-006-9008-0
- Kenneweg, S., Jackson, B. B., Deigmoeller, J., Eggert, J., & Cimiano, P. (2024, May). An empirical study on vague deictic temporal adverbials. In *Proceedings of the workshop on cognitive aspects of the lexicon @ lrec-coling 2024* (pp. 26–31). Torino, Italia: ELRA and ICCL.
- Lassiter, D., & Goodman, N. (2014, 01). Context, scale structure, and statistics in the interpretation of positive-form adjectives. *Proceedings of SALT*, 23, 587–610.
- Lassiter, D., & Goodman, N. D. (2017). Adjectival vagueness in a bayesian model of interpretation. *Synthese*, 194(10), 3801–3836. doi: 10.1007/s11229-015-0786-1
- Lawry, J., & Tang, Y. (2009). Uncertainty modelling for vague concepts: A prototype theory approach. *Artificial Intelligence*, 173(18), 1539-1558. doi: <https://doi.org/10.1016/j.artint.2009.07.006>
- Qing, C., & Franke, M. (2014, 06). Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. In (Vol. 24). doi: 10.3765/salt.v24i0.2412
- Schmidt, L., Goodman, N., Barner, D., Edu, B., & Tenenbaum, J. (2010, 09). How tall is tall? compositionality, statistics, and gradable adjectives.
- Solt, S., & Gotzner, N. (2012, 09). Experimenting with degree. In (Vol. 22). doi: 10.3765/salt.v0i0.2636
- Van Jaarsveld, H. J., & Schreuder, R. (1985). Implicit quantification of temporal adverbials. *Journal of Semantics*, 4(4), 327-339. doi: 10.1093/jos/4.4.327
- Williamson, T. (1998, 10). Vagueness. *The Philosophical Review*, 107. doi: 10.2307/2998388
- Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353. doi: [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)