**A Cognitive-statistical Analysis of Noun-based Time Adverbials in Biblical Hebrew**

Cody Kingham

University of Cambridge

cak47@cam.ac.uk

**Abstract**

This article utilizes a cognitive-statistical method to analyze the composition of over 3,400 Biblical Hebrew (BH) phrases marked for adverbial time function. Adverbials in general, and time adverbials in particular, have received relatively little treatment in the literature. A cognitive-statistical method attempts to trace statistical relations in language back to cognitive relations. The method works from the hypothesis that cognitive categories produce constructional co-occurrences. Time adverbials in BH are typically noun- or adverb-based. But what exactly is the difference between a 'noun' and 'adverb'? And is there such a thing as something in between? This article utilizes the cognitive-statistical method to examine how noun phrases are incorporated and transformed within the time adverbial function. This is done through two main experiments. The first details a statistical method to model parts of speech as tendencies along a noun-adverb continuum. An unsupervised clustering algorithm (PCA) is used to group words based on their co-occurrence frequencies with noun modifiers. The experiment confirms a broad division between nouns and adverbs but also provides evidence for intermediate categories. Second, a comparison is made between the modifiers of noun-based time adverbials and five other major arguments: adverbial location, (other) adjunct, complement, object, and subject. A statistical significance test (ΔP) is applied to 50,395 phrases to examine how the various arguments predict certain noun modifiers more strongly than others. The analysis finds that nominal time adverbials are strong predictors for definite article modification. This may be due to the need to more frequently anchor nominal time words with to a non-deictic, predicative reference point. These findings shed light on the syntax and semantics of time adverbials while opening up new avenues for exploring linguistic categories with statistics.

**A Cognitive-Statistical Approach**

With the exception of a few small studies, phrasal time adverbials in Biblical Hebrew have received little systematic, linguistic treatment.[[1]](#footnote-1) The reason for this may be because simple phrase adverbials present no obvious challenges for the classical linguistic method. As the grammars reiterate, time adverbials are simply adverbs or noun phrases that indicate optional time reference.[[2]](#footnote-2) But this simple description leaves many unanswered questions. For instance, what qualifies as an adverb or a noun? How do adverb-headed time phrases differ semantically from noun-headed phrases? How are noun phrases utilized within the time adverbial function?[[3]](#footnote-3)

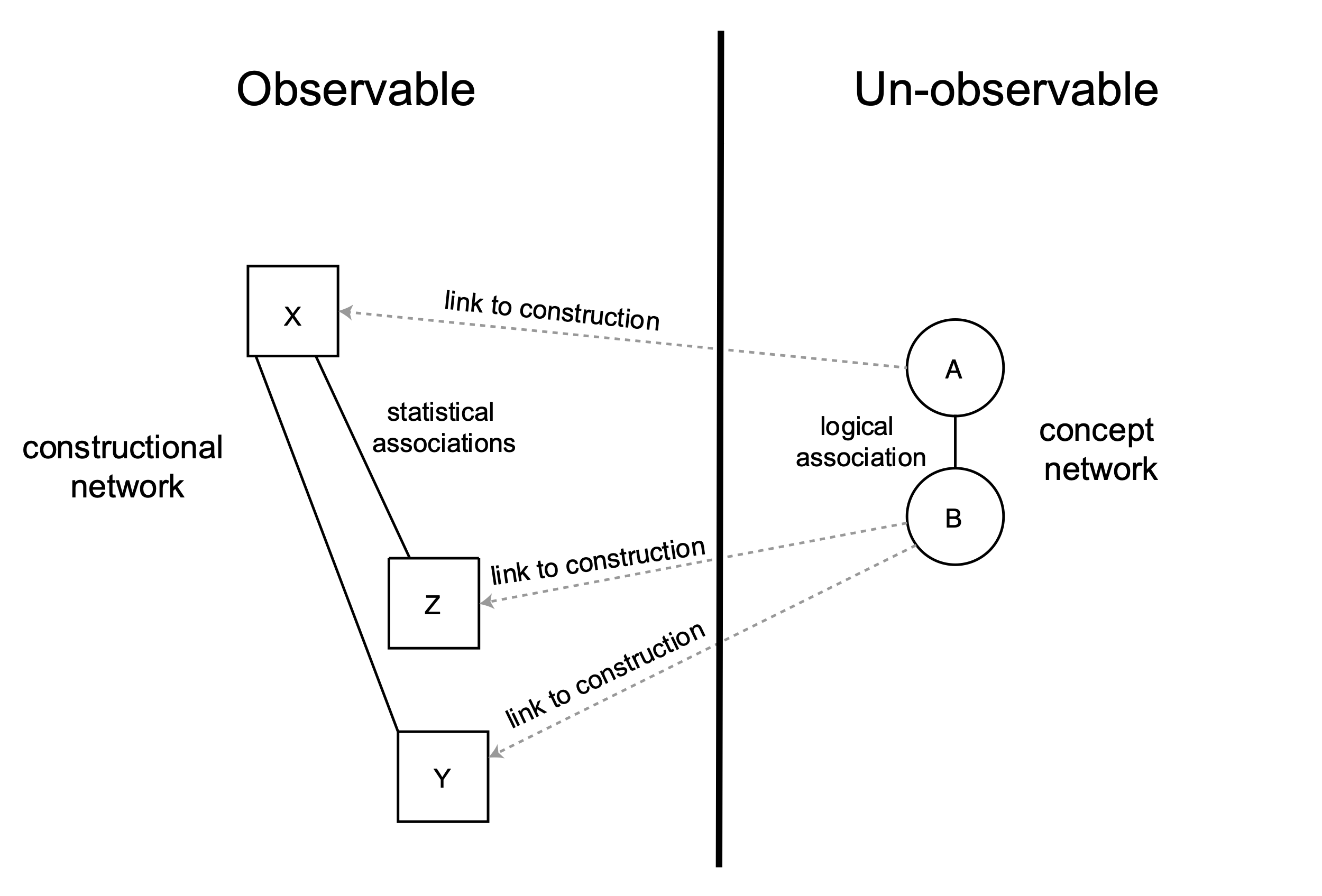
These questions are not trivial but go to the very foundation of grammatical inquiry. In recent years, Hebrew linguists have become more interested in foundational questions.[[4]](#footnote-4) The interest parallels a sea change in general linguistics, largely driven by new challenges from the cognitive linguistic perspective.[[5]](#footnote-5) At the same time, recent advancements in machine learning and brain imaging seem to confirm that language categories are learned rather than innate.[[6]](#footnote-6) This calls into question the degree to which *a priori* categories can be assumed within individual languages.[[7]](#footnote-7) In BH, linguists have long recognized that expected categories such as 'noun', 'adjective', and 'adverb' do not always fit.[[8]](#footnote-8) Given the new advancements, BH categories like 'noun' and 'adverb' are ripe for fresh evaluation.

The cognitive approach provides an elegant and powerful explanation for the huge variety of forms observed across and within world languages. The simple human impetus to form categories around observed data drives the creation of linguistic patterns. Language, then, is a cognitive tool for categorizing the world.[[9]](#footnote-9) Humans connect categories built from experiential knowledge to linguistic sounds and signs (cf. Saussure).[[10]](#footnote-10) The concepts of prototypes ("best example"), family resemblance, and gradient categories are central to the structure of categories.[[11]](#footnote-11) As it turns out, these structures line up with quantitative patterns observed in the world. Long-tailed distributions (e.g. Zipf's Law) pervade natural, biological, and sociological systems and may drive the formation of prototypes.[[12]](#footnote-12) Hebraists will recognize this distribution from lexeme frequencies: a handful of terms (e.g. **וְ** 'and', **הַ** 'the', **אָמַר** 'say') occur very frequently, whereas thousands of terms occur only once (*hapax legomena*). Language learners likely use the most frequent items, the prototypes, as reference points for analogically learning new categories.[[13]](#footnote-13)

The link between statistical patterns and cognition opens the door to an empirical method for studying language categories.[[14]](#footnote-14) Probabilistic patterns in the world provide humans with training data, which they then re-externalize using language.[[15]](#footnote-15) Mental associations between cognitive concepts thus give rise to statistical associations between language patterns.[[16]](#footnote-16) These associations appear most clearly in the phenomenon of cooccurrence, or collocation, between related constructions.[[17]](#footnote-17) For instance, **מֶלֶךְ** 'king' is statistically associated with **צִוָּה** 'command' in the Hebrew Bible (HB).[[18]](#footnote-18) Conceptually this is due to the relatedness of the two terms with 'authority'. Synonym relations can be identified by indexing which terms collocate with the same kinds of items; for example: **שַׂר** 'prince'. This statistical concept has been successfully extended to grammatical constructions, further demonstrating that lexicon and syntax are not fundamentally distinct.[[19]](#footnote-19) The collocation principle underlies a number of recent advancements in machine translation.[[20]](#footnote-20)

The cognitive-statistical approach attempts to work backward from statistical associations towards hypothetical cognitive associations, as illustrated below.

Figure 1: Mutual relationship between linguistic, cognitive, and statistical data



While linguists have some introspective access to cognitive categories, it is unclear how reliable our own judgments can be.[[21]](#footnote-21) Linguists should therefore seek to operationalize their intuitions into testable hypotheses. This approach provides cognitive linguists with a more objective means for testing their hypotheses.[[22]](#footnote-22)

In BH, Forbes has pioneered data-driven models to study word classes, using unsupervised clustering algorithms to automatically group words based on their co-occurrences.[[23]](#footnote-23) The approach yields promising results, with clear tendencies and recognizable groups. The methods of analysis for this study are different, but similarly reliant on the principle of collocation.

**Dataset and Methodology**

The objective of this analysis is to measure the makeup and collocational tendencies of time adverbial phrases while comparing them with other argument types. To that end, the study requires a dataset that has wide coverage of the relevant syntactic categories and is freely accessible for scholarly scrutiny. Data-driven research is not theory neutral, nor is it immune to bias. This is equally true in the natural sciences.[[24]](#footnote-24) Rather, scientific research aims to make its assumptions clear, gather and measure data methodically, test the assumptions against the data, and make the results available for public scrutiny.[[25]](#footnote-25) Any data-driven study that does not make its full dataset available does little to advance the field.

The open-sourced BHSA of the Eep Talstra Centre for Bible and Computer (ETCBC) meets the proposed criteria.[[26]](#footnote-26) The BHSA is accessed using Python and a corpus analysis package, Text-Fabric.[[27]](#footnote-27) It contains annotations for phrases interpreted to indicate adverbial time modification as well as numerous other argument roles, with coverage for the whole of BHS. The ETCBC built the annotations over a 40+ year history.[[28]](#footnote-28) As with any dataset, BHSA represents only one interpretation. It also possesses a few shortcomings. The analysis is only based on *qere* forms. Furthermore, relations below the phrase level are not always reliable, and there exists no tagging to explicitly indicate headship. For this reason, phrases must be pre-processed using a custom-built parser to isolate the necessary relations.[[29]](#footnote-29)

It is important to note that large scale corpus studies such as this require a more forbearing accuracy standard than traditional analyses. The retrieval and tagging of large quantities of text always assumes a tradeoff between precision and recall, i.e. a ratio of good or missed matches.[[30]](#footnote-30) Realistically, large-scale studies require a compromise between perfection (technically impossible) and coverage.[[31]](#footnote-31) Especially with the parsing of internal phrase relations, edge cases are likely to remain in this study's sample. This emphasizes the need for good exploratory data analysis. Results should not be treated as a black box but carefully examined in connection with the actual underlying data. Additionally, reliance on statistical significance helps avoid erroneous conclusions. The datasets, along with all the code for the analysis, are published online for scrutiny.[[32]](#footnote-32)

The primary dataset is a subset of all phrases in the Hebrew Bible stored in the BHSA. There are 99,426 phrases labeled as time adjunct, location adjunct, other adjunct, complement, object, or subject (henceforth Time, Loca, Adju, Cmpl, Objc, Subj). These phrases were tagged with human oversight using a computer-assisted method.[[33]](#footnote-33) Many of these phrases contain embedded sub-phrases that are not always well-marked. To ensure accuracy, the set is pruned down to a group of shorter, simple phrases. The resulting dataset contains 73,120 phrases that are parsed for head words and modifiers using a generally semantic definition.[[34]](#footnote-34) A test to see whether the subsampled phrases negatively affected any particular book's representation showed no significant effects.[[35]](#footnote-35)

Table 1: Phrase function frequencies and proportion of selected samples versus BHSA[[36]](#footnote-36)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Adju | Cmpl[[37]](#footnote-37) | Loca | Objc | Subj | Time |
| *BHSA freq.* | 9403 | 29568 | 2597 | 22394 | 31423 | 4041 |
| *sample freq.* | 6737 | 17348 | 2125 | 16496 | 26156 | 3442 |
| *sample proportion* | 0.72 | 0.59 | 0.82 | 0.74 | 0.83 | 0.85 |

Throughout the study, I will refer to the phrase labels such as Time, Subj, etc. to distinguish the samples from the actual linguistic categories. Thus, I do not assume a one-to-one correspondence between the two. The tagged phrases can be viewed as tools of analysis that help to approach the object of study. This is on analogy with other scientific objects of measurement (e.g. a ruler, or a scale) which approximate rather than exhaust reality.[[38]](#footnote-38) Alternatively the phrases and tags can be viewed as a hypothesis (one interpretation) in need of testing.

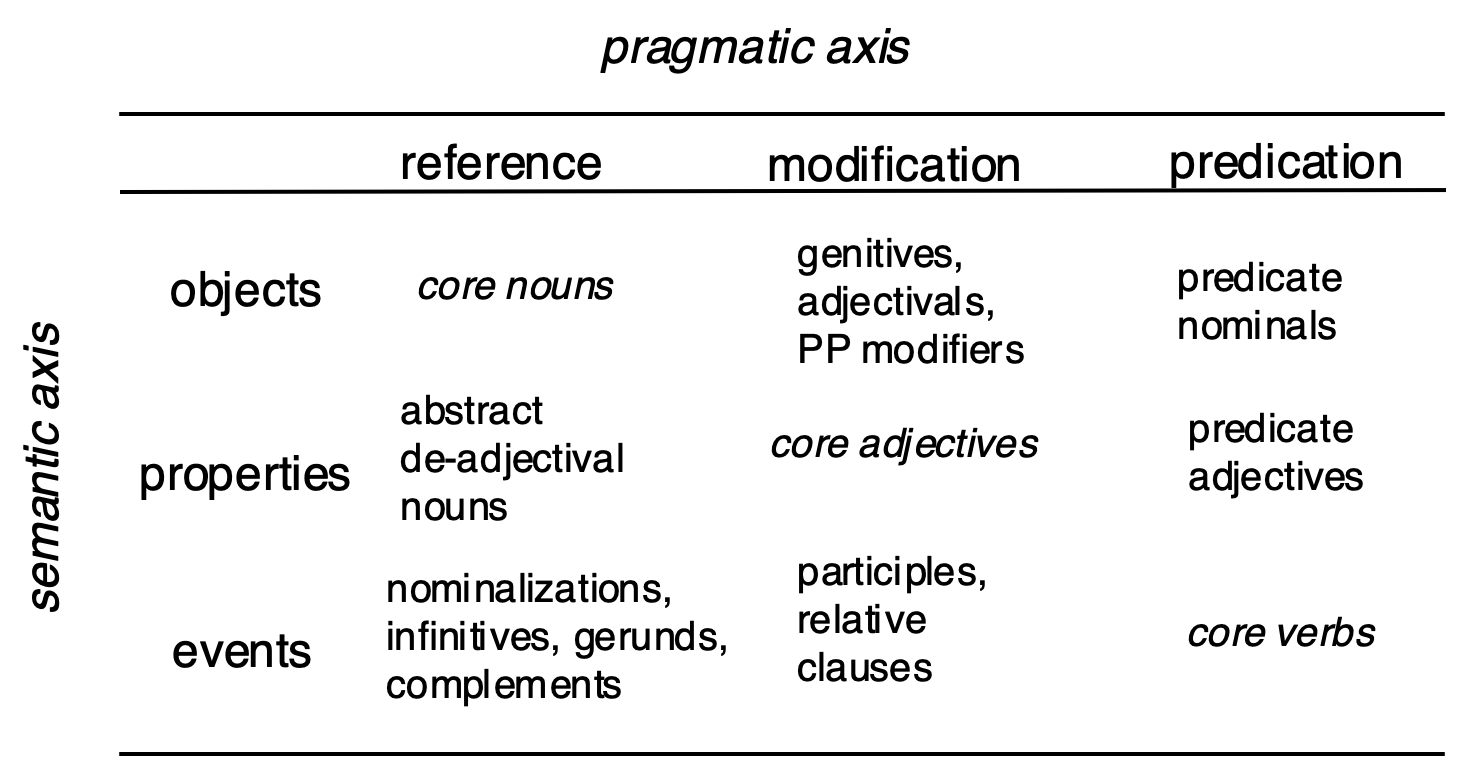
**Parts of Speech Experiment**

BH grammars tend to treat time adverbials under three different rubrics: adverbs, adverbial accusatives, and temporal prepositional phrases.[[39]](#footnote-39) Adverbial accusatives and prepositional phrases are typically noun-based. Yet the distinction between an adverb and noun in BH is not always clear.[[40]](#footnote-40) BH lacks a productive morphology for adverbs, unlike languages such as English which possesses morphemes like -*ly*.[[41]](#footnote-41) But even English and many languages do not always clearly distinguish between nouns and adverbs.[[42]](#footnote-42) One reason for this may be that adverbial function is already an abstract, peripheral role compared with other sentence arguments.[[43]](#footnote-43) As a result, adverbs tend to be sourced from a variety of word classes, resulting in a heterogenous makeup.[[44]](#footnote-44) In BH, the most common source for adverbs is nouns and adjectives.[[45]](#footnote-45)

The absence of a neat and tidy word class is not unique to adverbs but is endemic to parts of speech in languages around the world. Some languages, for instance, lack a clear morphological distinction between nouns and verbs.[[46]](#footnote-46) These problems have led cognitive linguists to abandon universal, language-inherent parts of speech.[[47]](#footnote-47) Instead, it is proposed that only semantic meaning is ubiquitous; that is, cognitive concepts which are universal to human experience can be mapped onto an endless range of language patterns.[[48]](#footnote-48)

Croft, in particular, has advocated a geometric approach to parts of speech, whereby various cognitive concepts are visualized in a conceptual space.[[49]](#footnote-49) He recognizes two main semantic axes in this space by which humans tend to organize their knowledge of the world: a semantic axis and a pragmatic axis:

Figure 2: Croft's two-dimensional parts of speech space[[50]](#footnote-50)



Constructions in language can point to any region or stretch within this space. For example, the English pattern -*ing* spans the events with reference (e.g. "running") and events with modification regions (e.g. "running man").[[51]](#footnote-51) In addition to the three main pragmatic acts, Croft provides a role which serves to "situate in a physical dimension," which includes locative time adverbials.[[52]](#footnote-52)

Of particular interest to this study is the way in which the adverbial role "recruits" words associated with the noun sense. Prototypical adverbs are words which have their temporal reference built-in, so to speak; they are typically deictic[[53]](#footnote-53) and sometimes have tense (e.g. **מָחָר** 'tomorrow').[[54]](#footnote-54) But adverbials in BH (and others) also utilize nouns to indicate location or duration in time. They can therefore be construed simultaneously as conceptual objects with referentiality (i.e. nouns) and modifiers indicating temporal location (adverbs).[[55]](#footnote-55) But what happens when a noun begins to increasingly "forget" its nominal properties and take on the life of an adverb? This possibility would imply that there exists a spectrum of words that exhibit more or less nominal behavior. Is there a way to test this hypothesis?

This section provides such an experiment to model nominal behavior in words that head Time phrases. The experiment will classify words based on how frequently they collocate with nominal modifiers. Following Croft, nouns in their most basic sense represent objects within a real or imagined physical space.[[56]](#footnote-56) As such, nouns possess various properties which relate their position and quantity within the space. Modifiers like definite articles and demonstratives situate nouns in the space (e.g. 'near' versus 'far'); possessives create relations with other objects; and numbers or quantifiers "select" the nouns.[[57]](#footnote-57)

For this experiment, nine types of noun modifiers are tagged in the sample when found with a phrase head: plural endings, pronominal suffixes, definite articles,[[58]](#footnote-58) construct relations (called 'genitive' here[[59]](#footnote-59)), demonstratives or ordinals, cardinal quantifiers, and other quantifiers (e.g. **כֹל**).[[60]](#footnote-60) Another label, Ø, indicates no modifying specification of the headword except for the allowance of prepositions. The tagged features and their respective quantities for Time are shown below; features are allowed to co-occur.

Table 3: Collocational frequencies between Time head and nominal modifiers

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cardinal | definite | demonstrative | genitive | ordinal | plural | quantifier | suffix | Ø |
| 411 | 1331 | 493 | 426 | 162 | 569 | 212 | 103 | 1148 |

A separate table is compiled which tabulates how frequently a given lexeme occurs with a given modifier. Words with a sample size <5 are dropped to ensure enough data is present for reliable analysis. The resulting table contains 41 lexemes. The counts for every lexeme × modifier are normalized to the total number of modifiers observed with a lexeme.[[61]](#footnote-61) The result is the proportion a given modifier represents of all a head's modifiers (read as a percentage when multiplied by 100).

Table 4: Excerpt of Time head × modifier co-occurrence matrix (original: 41×9)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| head | cardinal | definite | demonstrative | genitive | ordinal | plural | quantifier | suffix | Ø |
| יֹום | 165 | 819 | 383 | 279 | 108 | 384 | 171 | 41 | 9 |
| שָׁנָה | 188 | 42 | 7 | 22 | 25 | 82 | 4 | 2 | 0 |
| עַתָּה | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 354 |

Table 5: Excerpt of Time head × modifier co-occurrence matrix (original: 41×9)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| head | cardinal | definite | demonstrative | genitive | ordinal | plural | quantifier | suffix | Ø |
| יֹום | 0.07 | 0.35 | 0.16 | 0.12 | 0.05 | 0.16 | 0.07 | 0.02 | 0.0 |
| שָׁנָה | 0.51 | 0.11 | 0.02 | 0.06 | 0.07 | 0.22 | 0.01 | 0.01 | 0.0 |
| עַתָּה | 0.0 | 0.0 | 0.01 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.99 |

This multi-dimensional dataset would be tedious and error-prone to compare by hand (41×9 = 369 data points). Instead, an unsupervised clustering algorithm called Principle Component Analysis (PCA) is used to simultaneously compare all of the values and group them based on prevailing trends.[[62]](#footnote-62) The resulting graphs (Figure 6) should be interpreted with respect to space, where closer samples are more related based on the input features. Parts of speech categories from the lexicon are added to the first graph to show how they correspond with the PCA classification.[[63]](#footnote-63) The second graph plots the same data but with the text of the heads, and with a slight adjustment of their positions for readability. The blue arrows in the plot show what has influenced each word's placement on the graph.[[64]](#footnote-64)

Figure 3: PCA analysis of head word tendencies

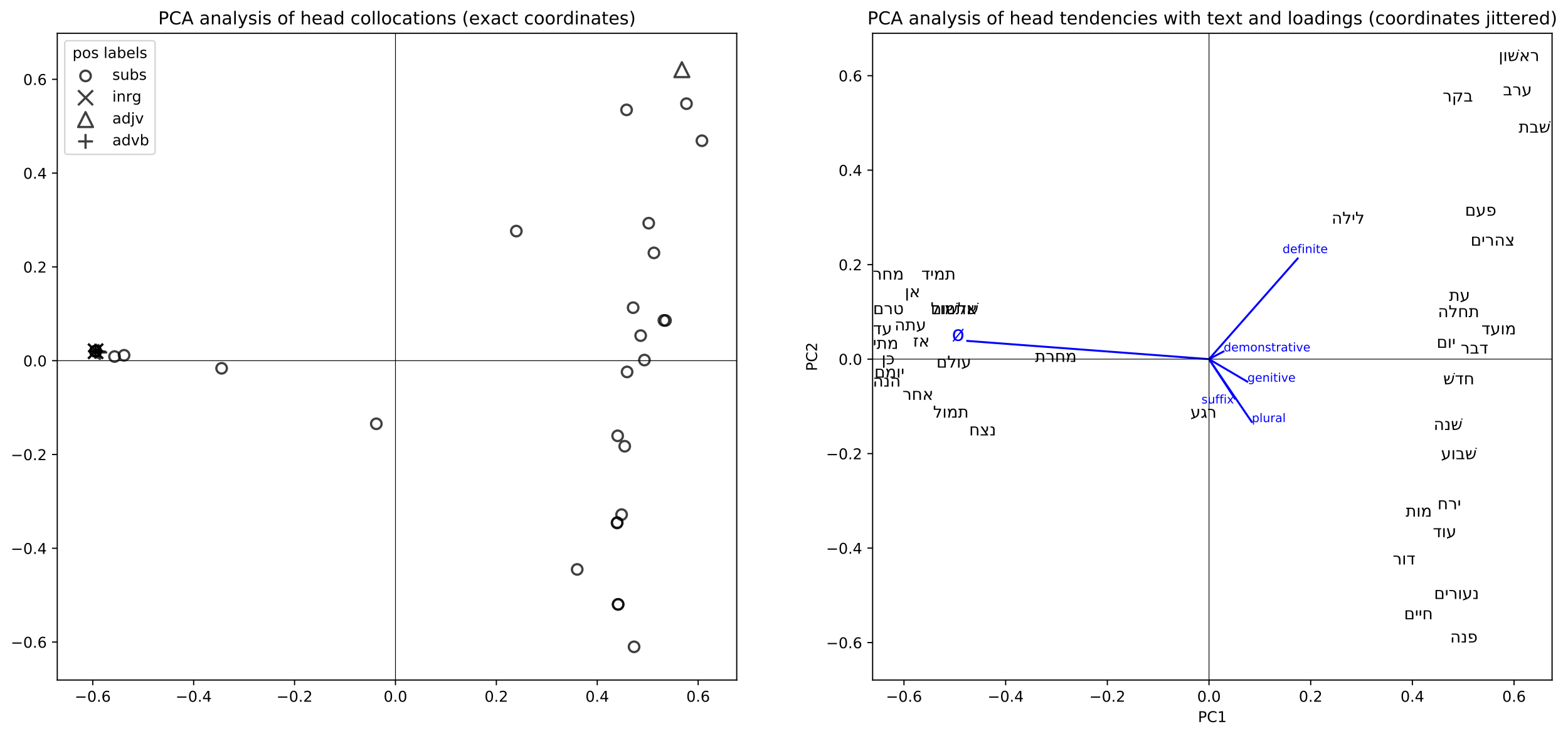
****

Table 6: Mean proportions for null (Ø) versus nominal modifiers by placement along x-axis

|  |  |  |
| --- | --- | --- |
|  | Ø | nominal |
| x<0 | 0.96 | 0.04 |
| x>0 | 0.04 | 0.96 |

The PCA graph reveals a clear separation of adverbs and nouns along a gradient. Table 5 shows that terms to the right of 0 occur with a nominal modifier on average 96% of the time, and on the left of 0 with null modification also with 96% on average. This reflects the strong tendencies seen visually in the graph. In cognitive terms, the items on the extreme ends of the plot represent potential prototypes. The separation confirms the intuition that Time adverbial heads exhibit a broadly binary tendency between nouns and adverbs. Null (Ø) modification is the strongest contributor to the separation as seen by the long blue arrow.[[65]](#footnote-65)

The noun side of the graph shows two subcategories separated by different modifier tendencies. The blue influences (loadings) such as definite + demonstrative, and genitive + plural + suffix demonstrate how PCA identifies features which "cooperate" together in linguistic expressions. Definites and demonstratives frequently co-occur, for instance, in attributive adjunct expressions (e.g. **הַיוֹם הַזֶה**)[[66]](#footnote-66); plurals and suffixes co-occur in such constructions as **נְעוּרֵיךְ** 'your youth' or **חַיָּיו** 'his life'.[[67]](#footnote-67) It is noteworthy that these latter two terms, clustered along with **מוֹת** 'death' are thematically similar as life-events. This may explain their association with the suffix, which relates these terms to a personal entity.

While revealing clear prototypical behavior, the graph also shows fuzzy and intermediate tendencies. Firstly, it demonstrates the shortcoming of traditional part of speech values. Diverse lexicon tags of noun (subs), interrogative particles (inrg), and adverbs (advb) are found grouped together to the left of the y-axis. The right side is also mixed, containing one adjective (**רִאשׁוֹן** 'first').[[68]](#footnote-68) Secondly, several terms sit at intermediate positions along the x-axis, especially **רֶגַע** 'instant', **מָחֳרָת** 'next day', and **לַיְלָה** 'night'. Selected text examples for these terms will now be analyzed to understand their behavior.

The table below contains the full sample set for **רֶגַע**, a relatively small set (n=7). **רֶגַע** is placed slightly on the particle side of the graph.

Figure 4: Sampled sentences for *רֶגַע*

|  |  |  |  |
| --- | --- | --- | --- |
| reference | modifier | sentence | translation |
| *Num 17:10* | Ø | **וַאֲכַלֶּ֥ה אֹתָ֖ם כְּרָ֑גַע** | I will destroy them instantly. |
| *Isa 54:8* | Ø | **בְּשֶׁ֣צֶף קֶ֗צֶף הִסְתַּ֨רְתִּי פָנַ֥י רֶ֨גַע֙ מִמֵּ֔ךְ** | In a flood of anger I hid my face at an instant from you. |
| *Ps 73:19* | Ø | **אֵ֤יךְ הָי֣וּ לְשַׁמָּ֣ה כְרָ֑גַע** | How they are ruined instantly! |
| *Exod 33:5* | cardinal | **רֶ֧גַע אֶחָ֛ד אֶֽעֱלֶ֥ה בְקִרְבְּךָ֖** | At a single instant I will come up in your midst. |
| *Isa 26:20* | quantifier | **חֲבִ֥י כִמְעַט־רֶ֖גַע** | Hide instantly! [lit. "in a short instant"] |
| *Ezra 9:8* | quantifier | **כִּמְעַט־רֶגַע֩ הָיְתָ֨ה תְחִנָּ֜ה מֵאֵ֣ת׀ יְהוָ֣ה אֱלֹהֵ֗ינוּ** | In just a short instant our supplication is before YHWH our God. |
| *Isa 27:3* | plural | **לִרְגָעִ֖ים אַשְׁקֶ֑נָּה** | Continuously I watered it. |

3 of the text examples are used without any kind of nominal modifier, 3 employ quantifiers, and 1 has a plural ending. Given the small sample size, we should withhold strong judgments about this term, however a number of observations are pertinent. Even though this word has a slight majority of nominal modifiers (4/7), the algorithm has situated it on the adverb side due to the stronger influence of the null feature. Another reason is due to the weaker influence of quantifiers. This is not just an accidental artifact of the PCA analysis; the feature weights are determined across the whole dataset by the degree, and thus predictive power, of their variability. Furthermore, an abstract word like 'instant' does not seem to fit the semantics of a prototypical noun.[[69]](#footnote-69) The data thus tentatively indicate a term that sits somewhere on the border between a noun and adverb.

**מֳחָרָת** 'next day' appears 26 times in the sample. There are 20 cases of null modification and 6 cases of nominal modification. Note that 11 null cases not shown are identical to the first example with **וַיְהִי**.[[70]](#footnote-70)

Table 7: Sampled modifiers with *מָחֳרָת*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cardinal | definite | demonstrative | genitive | ordinal | plural | quantifier | suffix | Ø |
| 0 | 1 | 0 | 5 | 0 | 0 | 0 | 0 | 20 |

Table 8: Selected samples of *מָחֳרָת*

|  |  |  |  |
| --- | --- | --- | --- |
| reference | modifier | sentence | translation |
| *Gen 19:34* | Ø | **וַֽיְהִי֙ מִֽמָּחֳרָ֔ת** | It was the subsequent day. |
| *1 Sam 5:4* | Ø | **וַיַּשְׁכִּ֣מוּ בַבֹּקֶר֮ מִֽמָּחֳרָת֒** | He arose early on [the] next morning. |
| *1 Sam 5:3* | Ø | **וַיַּשְׁכִּ֤מוּ אַשְׁדֹּודִים֙ מִֽמָּחֳרָ֔ת** | The Ashdodites arose early [the] next day. |
| *Judg 6:38* | Ø | **וַיַּשְׁכֵּם֙ מִֽמָּחֳרָ֔ת** | He arose [the] next day. |
| *Num 17:6* | Ø | **וַיִּלֹּ֜נוּ כָּל־עֲדַ֤ת בְּנֵֽי־יִשְׂרָאֵל֙ מִֽמָּחֳרָ֔ת עַל־מֹשֶׁ֥ה וְעַֽל־אַהֲרֹ֖ן** | On [the] next day the whole assembly of Israel grumbled against Moses and Aaron. |
| *Lev 19:6* | Ø | **בְּיֹ֧ום זִבְחֲכֶ֛ם יֵאָכֵ֖ל וּמִֽמָּחֳרָ֑ת** | On the day of your sacrifice it shall be eaten, and on [the] next day. |
| *Exod 32:6* | Ø | **וַיַּשְׁכִּ֨ימוּ֙ מִֽמָּחֳרָ֔ת** | They arose on [the] next day. |
| *Exod 9:6* | Ø | **וַיַּ֨עַשׂ יְהוָ֜ה אֶת־הַדָּבָ֤ר הַזֶּה֙ מִֽמָּחֳרָ֔ת** | YHWH did this very thing on [the] next day. |
| *Josh 5:12* | Ø | **וַיִּשְׁבֹּ֨ת הַמָּ֜ן מִֽמָּחֳרָ֗ת בְּאָכְלָם֙ מֵעֲב֣וּר הָאָ֔רֶץ** | The manna ceased on [the] next day when they ate from the land's yield. |
| *Josh 5:11* | genitive | **וַיֹּ֨אכְל֜וּ מֵעֲב֥וּר הָאָ֛רֶץ מִמָּֽחֳרַ֥ת הַפֶּ֖סַח מַצֹּ֣ות וְקָל֑וּי בְּעֶ֖צֶם הַיֹּ֥ום הַזֶּֽה׃** | They ate unleavened bread and roasted meat from the land's yield on [the] day after the Passover, on that very day. |
| *Lev 23:16* | genitive | **עַ֣ד מִֽמָּחֳרַ֤ת הַשַּׁבָּת֙ הַשְּׁבִיעִ֔ת תִּסְפְּר֖וּ חֲמִשִּׁ֣ים יֹ֑ום** | Up to [the] day after the seventh Sabbath you shall count fifty days. |
| *Lev 23:11* | genitive | **מִֽמָּחֳרַת֙ הַשַּׁבָּ֔ת יְנִיפֶ֖נּוּ הַכֹּהֵֽן׃** | On [the] day after the Sabbath, the priest shall wave it. |
| *1 Sam 20:27* | genitive | **וַיְהִ֗י מִֽמָּחֳרַ֤ת הַחֹ֨דֶשׁ֙ הַשֵּׁנִ֔י** | It will be on [the] day after the second month. |
| *1 Chr 29:21* | genitive | **וַיַּעֲל֨וּ עֹלֹ֜ות לַיהוָ֗ה לְֽמָחֳרַת֮ הַיֹּ֣ום הַהוּא֒ פָּרִ֨ים אֶ֜לֶף אֵילִ֥ים אֶ֛לֶף כְּבָשִׂ֥ים אֶ֖לֶף וְנִסְכֵּיהֶ֑ם וּזְבָחִ֥ים לָרֹ֖ב לְכָל־יִשְׂרָאֵֽל׃** | On [the] day after that day they offered up as burnt offerings to YHWH a thousand bulls, a thousand rams, a thousand sheep, and their drink offerings, and their sacrifices in abundance for all of Israel. |
| *Jonah 4:7* | definite | **וַיְמַ֤ן הָֽאֱלֹהִים֙ תֹּולַ֔עַת בַּעֲלֹ֥ות הַשַּׁ֖חַר לַֽמָּחֳרָ֑ת** | God appointed a worm at the dawning of the next day. |

In 20 of 26 cases (77%), **מָחֳרָת** appears without nominal modification, reflecting adverb behavior similar to its lexical cousin **מָחָר** 'tomorrow' (100% Ø). There are 6 cases of explicit nominalization with 5 construct relations and 1 (vocalic) definite article. Specifically, the construct cases construe **מָחֳרָת** as a specific 'next day' after the Passover, Sabbath (2x), second month, and 'that day' (**הַיּוֹם הַהוּא**). The use with the (apparent[[71]](#footnote-71)) definite article, **לַמָּחֳרָת** 'at the next day' represents another nominalization of the term.

But perhaps unmodified cases of **מָחֳרָת** truly "are" just nouns used without modifiers? Despite having a good sample size of 26, the behavior of **מָחֳרָת** differs significantly from the other regular nouns, resulting in its PCA placement. Furthermore, the explicitly nominalized cases of **מָחֳרָת** seem to interfere with an ongoing fusion to **מִן** 'since'. In nearly all cases, **מָחֳרָת** occurs with **מִן**, which appears to be semantically bleached. The case in Lev 23:16 makes this clear with the double prepositions: **עַד מִמָּחֳרַת** 'until the day after'. Yet, in 2 of the 6 explicit nominal cases, **מָחֳרָת** appears without **מִן** (**הַיּוֹם הַהוּא** **לְמָחֳרַת** 'on the day after that day', 1 Chr 29:21; **לַמָחֳרַת** 'on the day after', Jonah 4:7). These 2 cases demonstrate that the noun construal may evidence a transitional state in the semantics **מָחֳרָת**.[[72]](#footnote-72) That is, as **מִמָּחֳרָת** grammaticalizes within time, it begins to specialize for adverbial function and lose its noun association.

לַיְלָה 'night' is the most frequent intermediate word in the graph, with 65% of its 131 sampled forms collocating with nominal modifiers. This is compared with an average nominal modification of 93% on the noun-side of the graph.

Table 9: Frequency and proportion of *לַיְלָה* with modification type

|  |  |  |
| --- | --- | --- |
|  | nominal | Ø |
| *frequency* | 85 | 46 |
| *proportion* | 0.65 | 0.35 |

Table 10: Sampled modifiers with *לַיְלָה* [[73]](#footnote-73)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cardinal | definite | demonstrative | genitive | ordinal | plural | quantifier | suffix | Ø |
| 2 | 81 | 19 | 0 | 0 | 5 | 19 | 0 | 46 |

Table 11: Selected samples of *לַיְלָה*

|  |  |  |  |
| --- | --- | --- | --- |
| reference | modifier(s) | sentence | translation |
| *Gen 40:5* | cardinal | **וַיַּֽחַלְמוּ֩ חֲלֹ֨ום שְׁנֵיהֶ֜ם אִ֤ישׁ חֲלֹמֹו֙ בְּלַ֣יְלָה אֶחָ֔ד** | The two of them each dreamed his own dream one night. |
| *Gen 41:11* | cardinal | **וַנַּֽחַלְמָ֥ה חֲלֹ֛ום בְּלַ֥יְלָה אֶחָ֖ד אֲנִ֣י וָה֑וּא** | He and I dreamed a dream one night. |
| *Gen 19:5* | definite | **אַיֵּ֧ה הָאֲנָשִׁ֛ים אֲשֶׁר־בָּ֥אוּ אֵלֶ֖יךָ הַלָּ֑יְלָה** | Where are the men who came to you tonight? |
| *Exod 14:20* | quantifier+def. | **וְלֹא־קָרַ֥ב זֶ֛ה אֶל־זֶ֖ה כָּל־הַלָּֽיְלָה׃** | Neither one approached the other for the whole night. |
| *Judg 16:3* | quantifier+def. | **וַיִּשְׁכַּ֣ב שִׁמְשֹׁון֮ עַד־חֲצִ֣י הַלַּיְלָה֒** | And Samson laid down until [the] middle of the night. |
| *Ps 121:6* | definite | **יֹומָ֗ם הַשֶּׁ֥מֶשׁ לֹֽא־יַכֶּ֗כָּה וְיָרֵ֥חַ בַּלָּֽיְלָה׃** | By day the sun will not strike you, nor the moon in the night. |
| *2 Chr 7:12* | definite | **וַיֵּרָ֧א יְהוָ֛ה אֶל־שְׁלֹמֹ֖ה בַּלָּ֑יְלָה** | YHWH appeared to Solomon in the night. |
| *Gen 19:35* | definite+demon. | **וַתַּשְׁקֶ֜יןָ גַּ֣ם בַּלַּ֧יְלָה הַה֛וּא אֶת־אֲבִיהֶ֖ן יָ֑יִן** | So also that night they gave their father wine to drink. |
| *2 Chr 1:7* | definite+demon. | **בַּלַּ֣יְלָה הַה֔וּא נִרְאָ֥ה אֱלֹהִ֖ים לִשְׁלֹמֹ֑ה** | On that night God appeared to Solomon. |
| *Isa 21:8* | qual. + def. + pl. | **וְעַל־מִ֨שְׁמַרְתִּ֔י אָנֹכִ֥י נִצָּ֖ב כָּל־הַלֵּילֹֽות׃** | I stand at my guard through all the nights. |
| *Song 3:1* | definite + plural | **עַל־מִשְׁכָּבִי֙ בַּלֵּילֹ֔ות בִּקַּ֕שְׁתִּי אֵ֥ת שֶׁאָהֲבָ֖ה נַפְשִׁ֑י** | Upon my bed in the nights I seek he whom my soul loves. |
| *Gen 14:15* | Ø | וַיֵּחָלֵ֨ק עֲלֵיהֶ֧ם׀ לַ֛יְלָה ה֥וּא וַעֲבָדָ֖יו | He and his servants were split up at night. |
| *Num 14:14* | Ø | וּבְעַמֻּ֣ד עָנָ֗ן אַתָּ֨ה הֹלֵ֤ךְ לִפְנֵיהֶם֙ יֹומָ֔ם  וּבְעַמּ֥וּד אֵ֖שׁ לָֽיְלָה׃ | In a pillar of cloud you walk before them by day and in a pillar of fire at night. |
| *Judg 9:34* | Ø | וַיָּ֧קָם אֲבִימֶ֛לֶךְ וְכָל־הָעָ֥ם אֲשֶׁר־עִמֹּ֖ו לָ֑יְלָה | Abimelek arose, with all the people with him, at night. |
| *2 Kgs 7:12* | Ø | וַיָּ֨קָם הַמֶּ֜לֶךְ לַ֗יְלָה | The king arose at night. |
| *Hos 4:5* | Ø | וְכָשַׁ֧ל גַּם־נָבִ֛יא עִמְּךָ֖ לָ֑יְלָה | Even the prophet will stumble with you at night. |
| *Neh 2:12* | Ø | וָאָק֣וּם׀ לַ֗יְלָה אֲנִי֮ וַאֲנָשִׁ֣ים׀ מְעַט֮ עִמִּי֒ | I arose at night along with the few men who were with me. |
| *2 Chr 35:14* | Ø | כִּ֤י הַכֹּהֲנִים֙ בְּנֵ֣י אַהֲרֹ֔ן בְּהַֽעֲלֹ֛ות הָעֹולָ֥ה וְהַחֲלָבִ֖ים  עַד־לָ֑יְלָה | For the priests, sons of Aaron, were offering up the burnt offering and the fat offering until night. |

Again there is the temptation to consider the null modifications of **לַיְלָה** as simple unmodified uses of a noun. Indeed, the PCA analysis does place **לַיְלָה** on the nominal side of the graph, though in an intermediate position. Yet other terms like **יוֹם** 'day', **בֹּקֶר** 'morning', or **עֶרֶב** 'evening' which are likewise well-represented in the dataset do not have near the amount of null-modified cases.[[74]](#footnote-74) **לַיְְלָה** clearly has a distinctive profile.

As with **מָחֳרָת**, there may be a diachronic reason for the mixed behavior of לַיְלָה tied up with adverbial specialization. Namely, it may be the case that the ָה at the end of לַיְלָה originates from the *heh locale* found in locative adverbs (e.g. **אַרְצָה** 'to the land', Is 8:23).[[75]](#footnote-75) The locative ending would be semantically extended as a time locator. Similar to **מָחֳרָת**, then, this may be a case of ongoing specialization towards the adverbial function which results in the loss of a prototypical noun sense.

Aside from intermediate words in the PCA plot, even those with decisive tendencies can be used in non-prototypical ways. **עוֹלָם** 'forever' occurs with Ø modification in 95.5% of its samples (170/178) with 8 exceptions:

Table 12: Selected nominalized samples of *עוֹלָם*

|  |  |  |  |
| --- | --- | --- | --- |
| reference | modifier | sentence | translation |
| *1 Kgs 8:13* | plural | **בָּנֹ֥ה בָנִ֛יתִי בֵּ֥ית זְבֻ֖ל לָ֑ךְ מָכֹ֥ון לְשִׁבְתְּךָ֖ עֹולָמִֽים׃** | I have surely built a lofty house for you, a place for your dwelling for eternity. |
| *Isa 45:17* | plural+genitive | **וְלֹא־תִכָּלְמ֖וּ עַד־עֹ֥ולְמֵי עַֽד׃ פ** | You will not be ashamed unto eternity on. |
| *Jer 28:8* | definite | **הַנְּבִיאִ֗ים אֲשֶׁ֨ר הָי֧וּ לְפָנַ֛י וּלְפָנֶ֖יךָ מִן־הָֽעֹולָ֑ם** | The prophets who were before me and you since eternity past... |
| *Joel 2:2* | definite | **כָּמֹ֗הוּ לֹ֤א נִֽהְיָה֙ מִן־הָ֣עֹולָ֔ם** | There has not been one like it since eternity past. |
| *Ps 61:5* | plural | **אָג֣וּרָה בְ֭אָהָלְךָ עֹולָמִ֑ים** | I will dwell in your tent for eternity. |
| *Ps 77:8* | plural | **הַֽ֭לְעֹולָמִים יִזְנַ֥ח׀ אֲדֹנָ֑י וְלֹֽא־יֹסִ֖יף לִרְצֹ֣ות עֹֽוד׃** | Shall my lord reject [me] for eternity and never be pleased again? |
| *1 Chr 17:14* | definite | **וְהַֽעֲמַדְתִּ֛יהוּ בְּבֵיתִ֥י וּבְמַלְכוּתִ֖י עַד־הָעֹולָ֑ם** | I will establish it in my house and kingdom unto eternity. |
| *2 Chr 6:2* | plural | **וַֽאֲנִ֛י בָּנִ֥יתִי בֵית־זְבֻ֖ל לָ֑ךְ וּמָכֹ֥ון לְשִׁבְתְּךָ֖ עֹולָמִֽים׃** | I built a lofty house for you, and a place for your dwelling for eternity. |

This case shows that even a term with prototypical adverb behavior can be semantically construed as a noun.[[76]](#footnote-76) Similar to **מָחֳרָת** with **מִן**, **עוֹלָם** appears formulaically with either **לְ** or **עַד** 'until' (157/178); 5 of the 18 cases without 'until' contain nominal modifiers. Given that **עוֹלָם** morphs into 'world' in Medieval Hebrew,[[77]](#footnote-77) this may show a shift toward nominalization, or it could represent a split in the word's semantics.[[78]](#footnote-78)

Even strongly prototypical nouns can be construed as abstract adverbs. **יוֹם** 'day' occurs with nominal modifiers in 1339/1351 (99%) of sampled cases. Yet the cases below demonstrate extensions of **יוֹם** into an adverb.

Table 13: Select adverbialized samples of *יוֹם*

|  |  |  |  |
| --- | --- | --- | --- |
| reference | modifier | sentence | translation |
| *Isa 43:13* | Ø | **גַּם־מִיֹּום֙ אֲנִ֣י ה֔וּא** | Surely since time past I am he. |
| *Ezek 48:35* | Ø | **וְשֵׁם־הָעִ֥יר מִיֹּ֖ום יְהוָ֥ה׀ שָֽׁמָּה׃** | And the name of city since time past shall be 'YHWH is there'. |

These examples closely resemble in meaning and form the construction **מֵעוֹלָם** 'from eternity'.[[79]](#footnote-79)

These data grant a window into the way conceptual categories are shaped and molded within the time adverbial function. Even concepts which have very real instantiations in the world such as 'night' or 'day' can undergo a semantic construal in service of an adverbial role. These results show that individual words themselves are not nouns, but rather words which are strongly associated with the noun concept. These associations can shift or change over time as words become increasingly grammaticalized with other constructions such as prepositions. Furthermore, these data show that phenomena like ellipsis cannot explain all of the aberrations from expected part of speech categories.[[80]](#footnote-80)

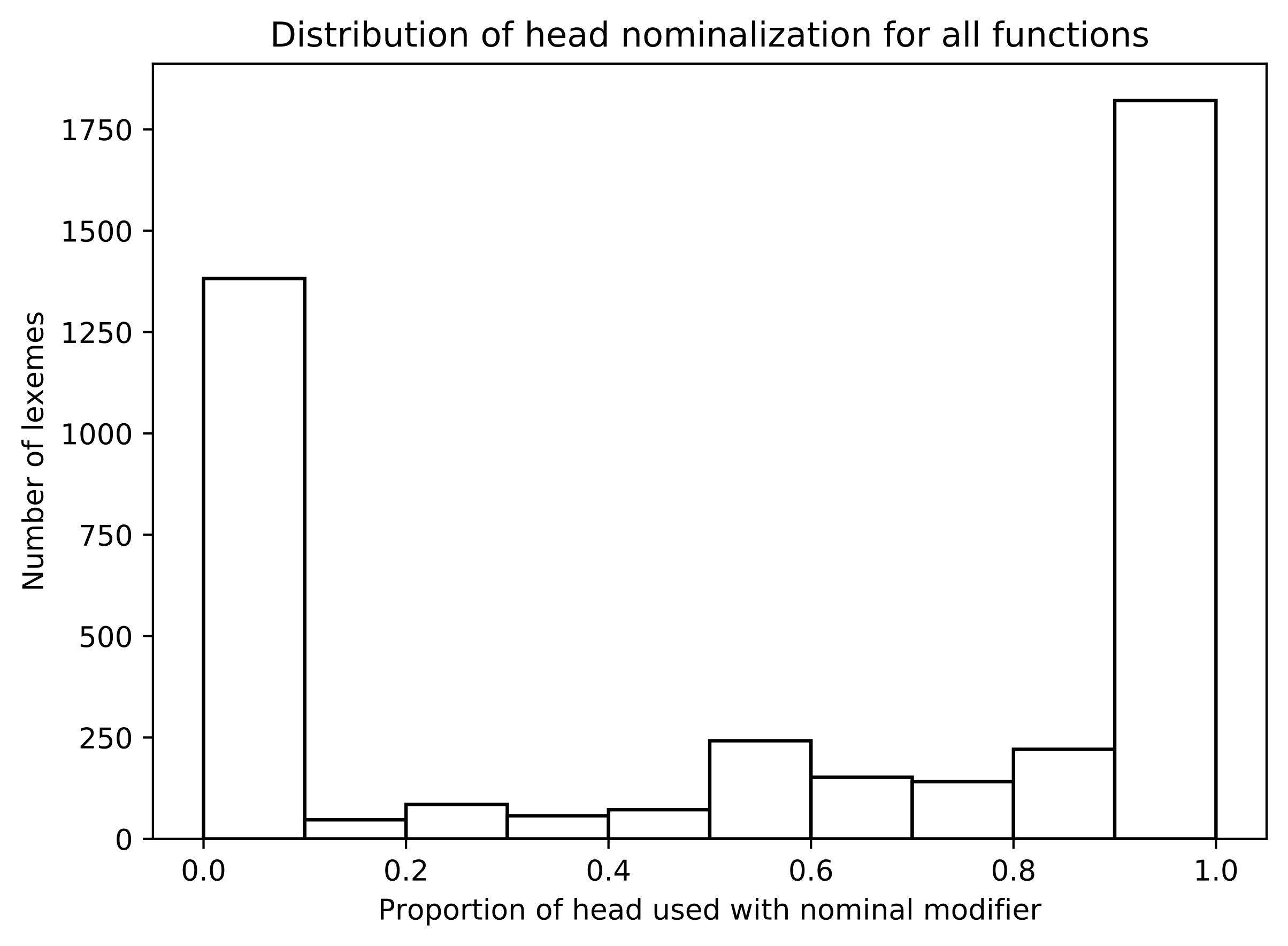
**Testing Time Function Predictions for Nominal Modifiers**

Up till now, we have examined the way various head modifiers can be used to distinguish adverb and nominal senses in time adverbials. The next question is how do noun modifiers uniquely contribute to the semantics of time? Plural morphemes, definite articles, constructs, etc. are sometimes portrayed in grammars as arbitrary, movable components of noun phrases. They are likened to a kind of Lego brick. However, quantitative methods reveal subtle differences in the way modifiers are deployed across various roles. This section will detail a comparison between phrases marked as Time and Loca, Adju, Cmpl, Objc, and Subj. Some roles predict more strongly certain modifiers. Modifiers thus resemble magnets more than blocks, with various attractions and repulsions to various arguments.

The comparison is accomplished with a statistical test called ΔP. ΔP has been used in psycholinguistics to formally test associative learning, that is, learning based on a supplied cue and a conditioned response.[[81]](#footnote-81) The more predictive a cue is for a response, the more associated the two categories. ΔP is a unidirectional measure, meaning association is measured with respect to the cue. It is also contingency-based, meaning that the strength of an association is measured over against the strength of other associations and the sample size.[[82]](#footnote-82) ΔP outputs a decimal ratio ranging from -1 to 1.[[83]](#footnote-83) -1 means the response is 100% less likely given the cue (i.e. response is repelled), whereas 1 represents 100% more likely (attracted).[[84]](#footnote-84) For this experiment, each of the 6 phrase functions are modeled as cues with 10 modifiers as responses; the analysis tests how strongly a given function predicts a given modifier.

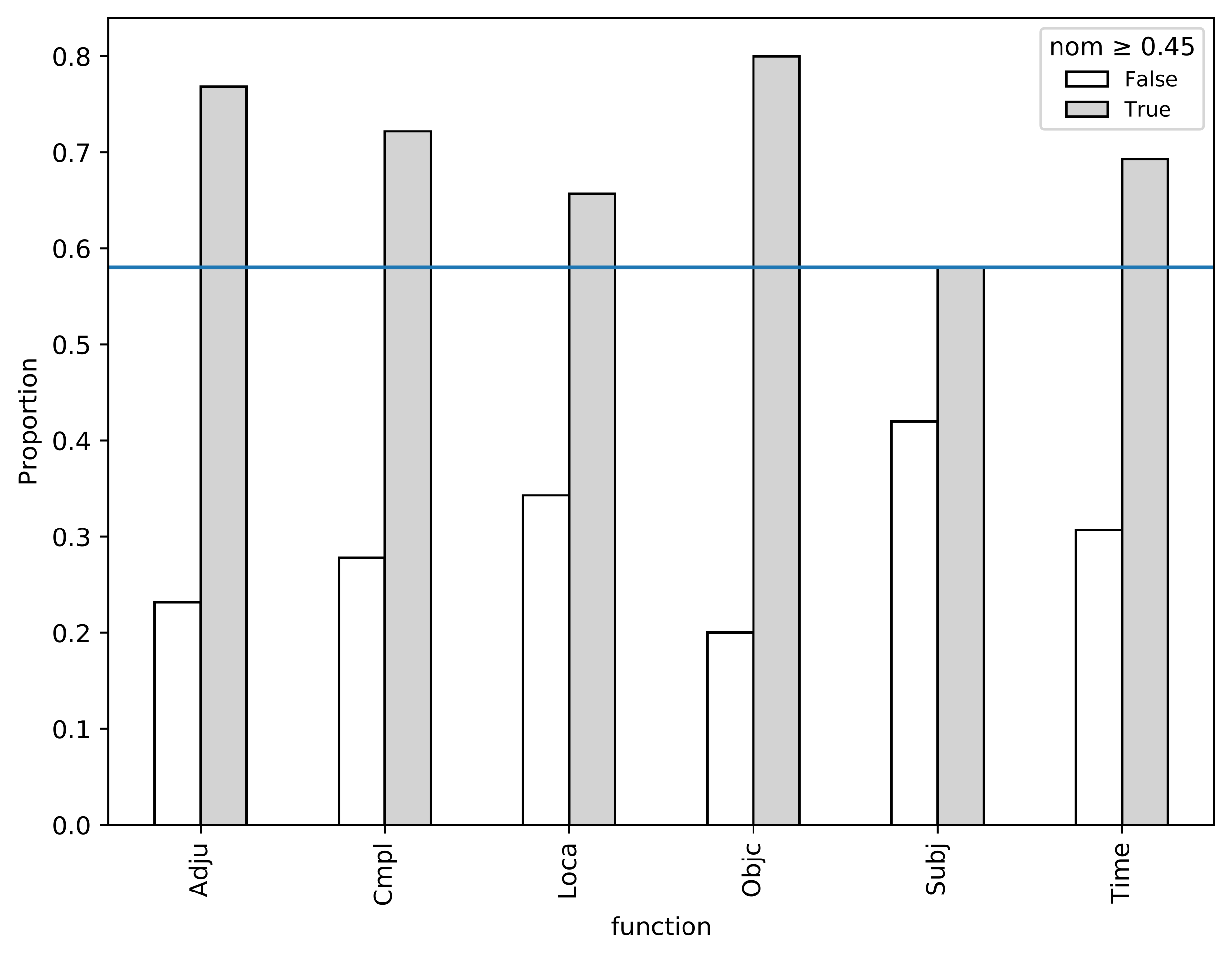
Before running the test, we first need to isolate a dataset of suitable noun-type samples across the 73,120 parsed phrases. This cannot be done manually, and as seen in the previous section words exhibit a range of behavior. Thus, lexicon parts of speech labels are insufficient. Instead, we will use the statistical behavior of head words to isolate a useful sample. To get an idea of the range of nominal behavior in the dataset, a count is made per every head lexeme of every instance where a nominal modifier occurs.[[85]](#footnote-85) The counts are then normalized to the total frequency of the lexeme in the initial sample. The result is a proportion of a given word's total instances of modification. The range of resulting values is illustrated in the plot below.

Figure 5: Head lexeme nominal modifier tendencies (N=73,120)



We see, for example, that >1,750 lexemes occur with between 90-100% nominal modification. This aligns with the broadly binary division observed in the part of speech experiment, with some ambiguity in the center. Earlier, modifier tendencies were measured with sample sizes of at least 5; whereas here there is no minimum. This appears to be why more terms straddle the middle.[[86]](#footnote-86) Given that the majority of nominal uses appear to occur with 50% and greater modification, I selected a cut-off point at 45%, giving a little flexibility. Thus, the sample includes phrases headed by those lexemes with a nominalization ratio of ≥0.45. The resulting dataset contains 50,395 samples (difference of 22,725). The effects of these selection criteria per function are illustrated in the graph below.

Figure 6: Proportions of nominal phrases by phrase function



Note that the Subj function is most heavily affected. This is due to a high frequency of proper names (e.g. **יְהֹוָה** N=2,169) and pronouns (e.g. **הוּא** N=904). The same is true of Loca (e.g. **שָׁם** N=309; יְרוּשָׁלִַם N=50, etc.).[[87]](#footnote-87) Despite these effects, all functions retain ≥58% of their occurrences, resulting in a large sample of nominal phrases.

For each phrase in the sample, a count of 10 modifier types are made, including those from the part of speech experiment plus a prepositional modifier. The counts are organized by function. Another table also tallies counts by genre, to control for various genre profiles.[[88]](#footnote-88) The table below contains raw counts for the whole sample.

Table 14: Modifier collocation frequencies for nominal phrases (all genres)[[89]](#footnote-89)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| function | cardinal | definite | demonstrative | genitive | ordinal | plural | preposition | quantifier | suffix | ø |
| *Adju* | 354 | 1175 | 76 | 1530 | 10 | 1372 | 4967 | 291 | 1395 | 652 |
| *Cmpl* | 568 | 3775 | 227 | 3532 | 9 | 3220 | 11877 | 431 | 3684 | 892 |
| *Loca* | 30 | 584 | 32 | 506 | 2 | 244 | 1277 | 69 | 196 | 58 |
| *Objc* | 628 | 1908 | 278 | 2873 | 25 | 4062 | 4249 | 599 | 4307 | 2392 |
| *Subj* | 631 | 3104 | 185 | 3759 | 25 | 5630 | 76 | 756 | 3804 | 1873 |
| *Time* | 411 | 1326 | 491 | 415 | 162 | 562 | 1578 | 209 | 102 | 101 |

The ΔP tests are run on the whole dataset, as well as per genre. The results are illustrated in Figure 6 and Figure 7. To aid interpretation, the resulting data tables are visualized as a heatmap. A heatmap color-codes the various intersecting datapoints with a shade of red, indicating higher attraction, or blue, indicating higher repulsion.[[90]](#footnote-90) The functions along the y-axis are the cues (predictors) whereas the modifiers on the x-axis are the responses.[[91]](#footnote-91) A PCA analysis similar to the one used to analyze Time heads is also shown, which clusters the functions based on their modifier predictions across all genres.

Figure 7: ΔP contingency tests applied to function head modifiers with PCA clustering

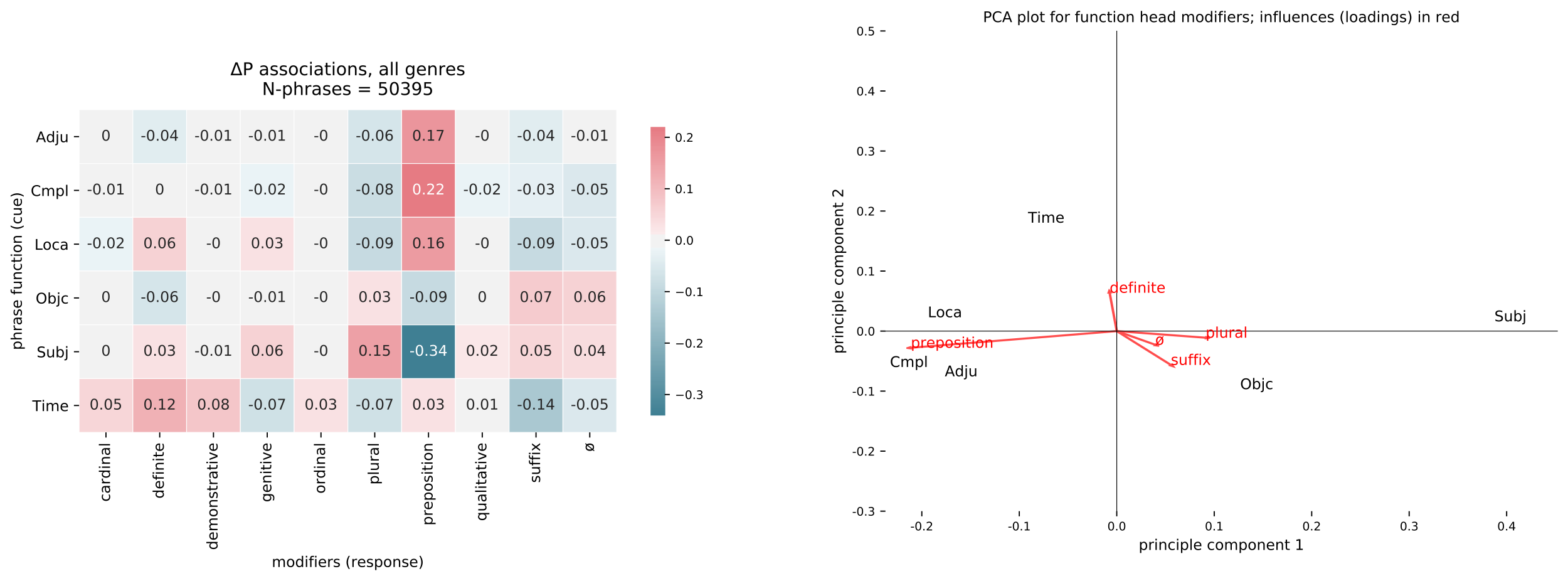
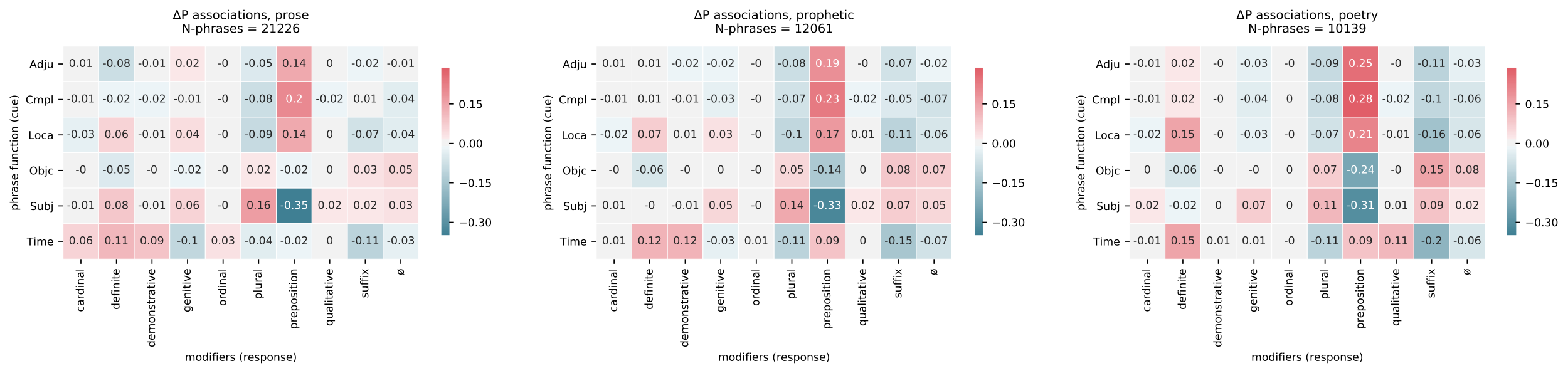


Figure 8: ΔP contingency tests by primary genres

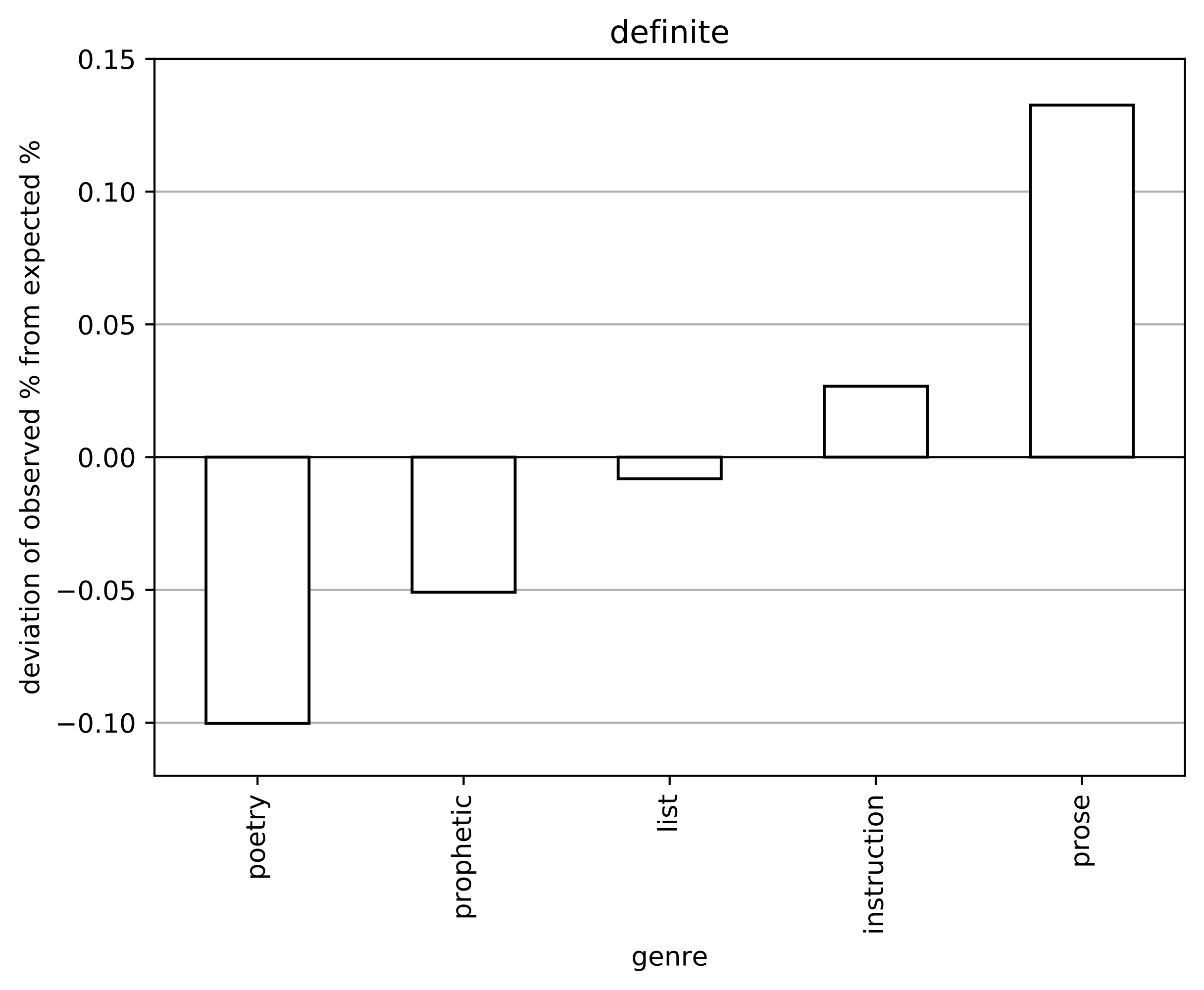
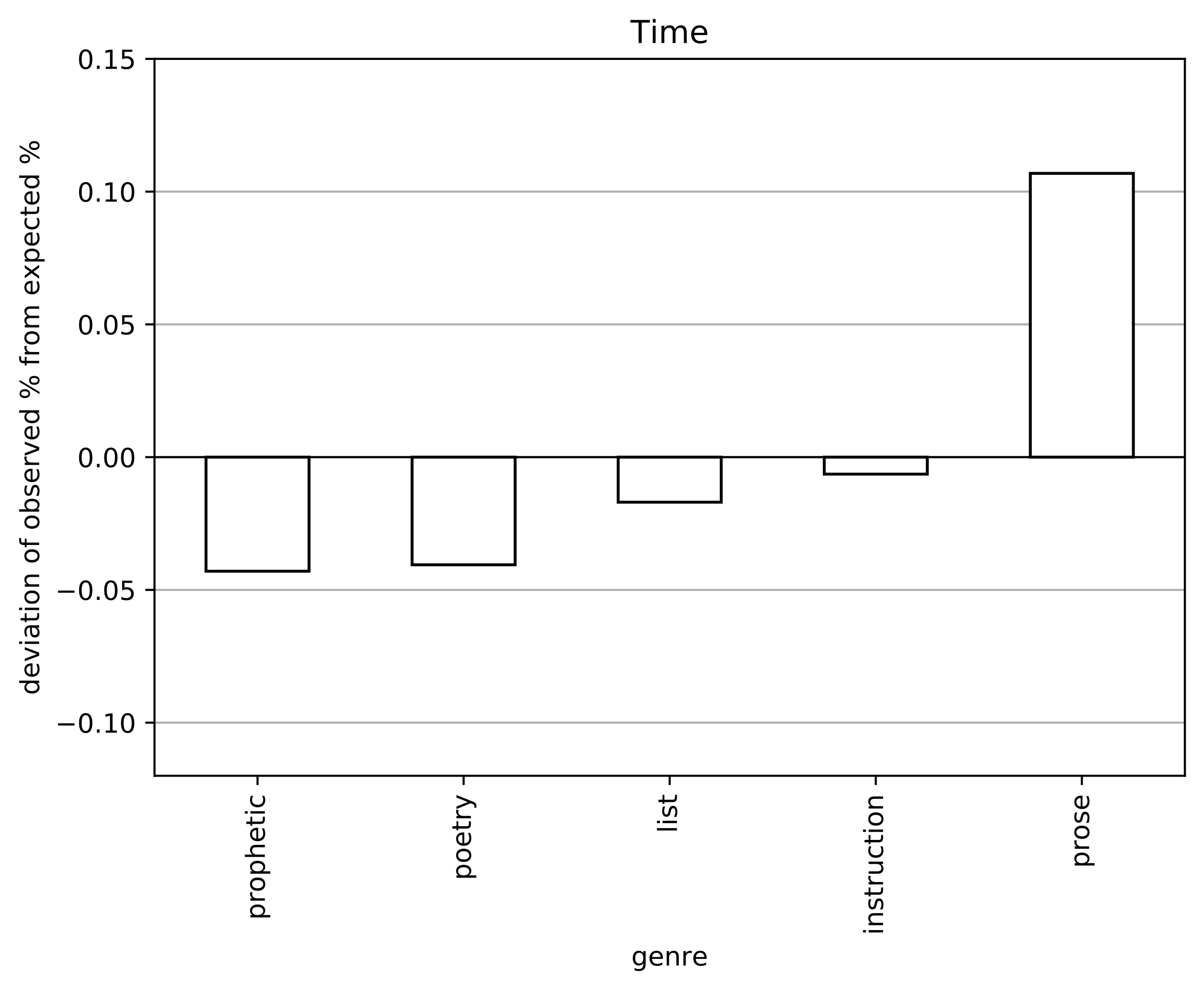


Several noteworthy trends can be seen for Time. For each genre, and in aggregate, Time is most predictive for definiteness (12% more likely across all genres). This is surprising, since this association is not intuitively associated with time expression. The reason for this relationship will be investigated further below. Time also has a mixed preference for prepositions; it is weak in prose (2% less likely [-]) though stronger in prophecy and poetry (9% more likely [+]). This medial position contributes to its slightly closer placement to Objc in the PCA plot. Time predicts demonstratives (+8%, e.g. **בַּיּוֹם הַהוּא** 'on that day'), cardinals (+5%, e.g. **ארְבָּעִים שָׁנָה** 'forty years'), and ordinals (+3%, e.g. **בַּחֹדֶשׁ הָרִשׁוֹן** 'on the first month'). Time has a negative prediction for suffix modification (-14%), plural endings (-7%), and genitive (construct) modifications (-7%). Time does occur with these modifiers, but they are relatively underrepresented; the Subj role, on the other hand, predicts the plural (+15%) as well as the suffix (+5%), a profile it shares with Objc.

What can these tendencies tell about time adverbial semantics? The analysis progresses down two trails. The first follows the intriguing higher preference for definiteness in Time. Rather than a simple coincidence, this tendency appears to reflect a semantic association between nominal time expression and non-deictic anchoring. The second path leads toward the weaker association with prepositions, a behavior predicted by the adverbial accusative lineage present in time indicators. The other attracted modifiers, i.e. cardinals, demonstratives, ordinals, are necessarily handled under these two lines of inquiry.

Time's strong prediction for definiteness (+12%) across all contexts is a profile it shares with Loca (+6%). Before any functional reasons can be considered for this relationship, it is necessary to first exclude other explanations. One possibility is that genre biases might cause a higher representation of definiteness. Both Time and definiteness have a clear genre bias in the sample. This can be quantified using a measure of expected proportional representation.[[92]](#footnote-92) For instance, prose accounts for 43% of all phrases in the sample. Thus, prose should be expected to account for 43% of any given function's uses if that function is evenly distributed. Any percentage above or below 43% represents a deviation from the expected value. A measure of deviation for Time and definiteness found that both were overrepresented in prose by 10% and 13%, respectively.

Figure 9: Deviation of observed proportions from expected proportions by genre for Time and definite

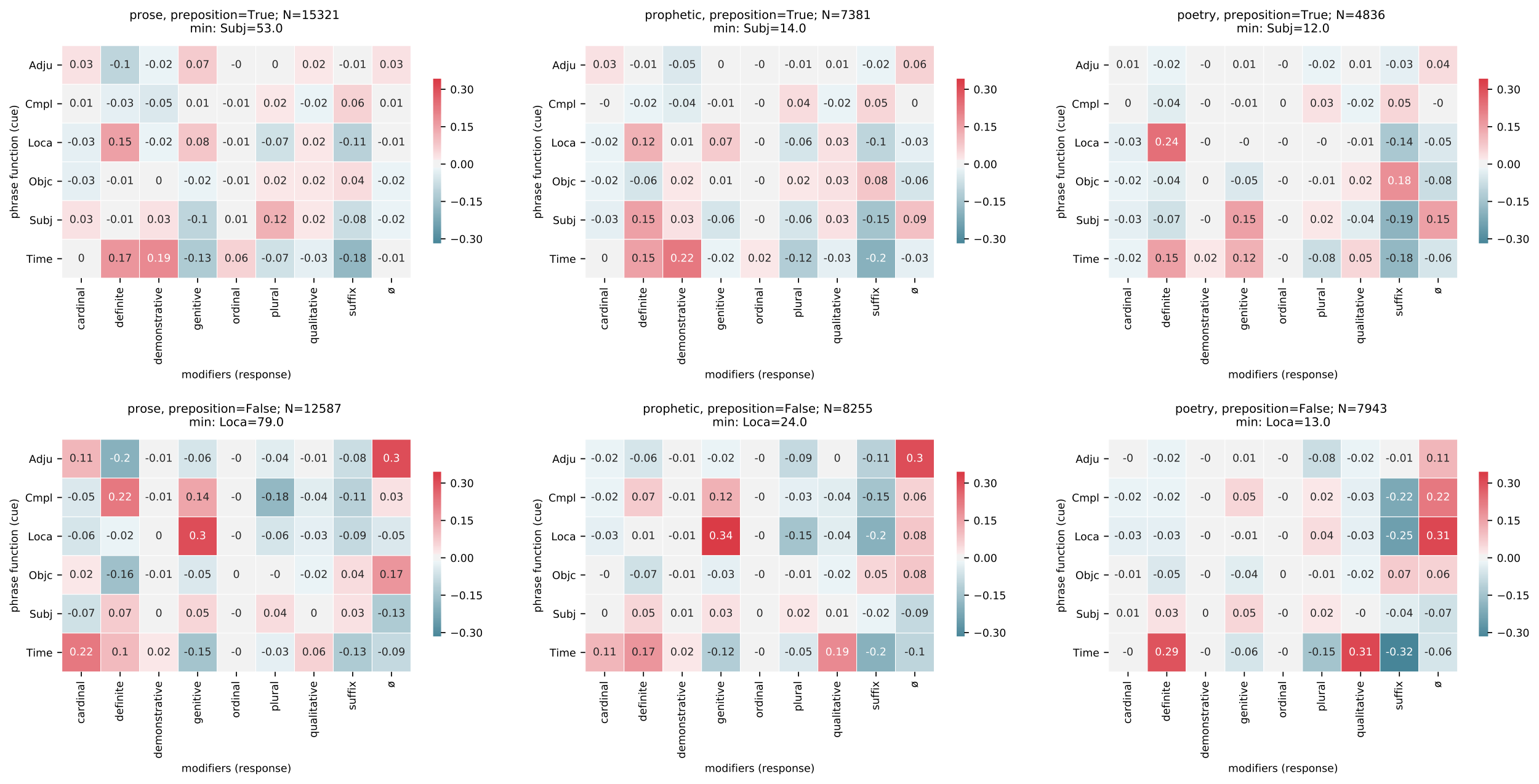


Yet, despite a genre bias, the data in Figure 7 (pg. 26) show that Time's prediction for definiteness actually increases in poetry over prose (+0.15 versus +0.11), despite a corresponding decrease in Subj and Objc. This means that the preference trends in the opposite direction of other phrase types. Thus, genre bias can be ruled out as a contributing factor for Time's definiteness preference.

Another possibility is that the higher degree of definiteness can be attributed to Tiberian interference. Gesenius argued that the ubiquitous use of a vocalized article with בְּ + יוֹם was evidence of "textual corruption" by the Masoretes.[[93]](#footnote-93) Ley similarly maintained that many vocalized articles in poetry should not be considered original.[[94]](#footnote-94) More recently, Bekins has noted similar suspicions.[[95]](#footnote-95) Barr takes a more cautious position, noting that a higher presence of vocalized articles in poetry seems to preclude Tiberian influence.[[96]](#footnote-96) Furthermore, their preference for "shorter words" like יָם 'sea' and טוֹב 'good', indicates the potential for either a diachronic or semantic explanation.[[97]](#footnote-97)

In order to examine the possibility of Tiberian influence on definiteness, another set of tests are run with and without the preposition across the three main genres (Figure 10). The additional stipulation of "with/without preposition" causes certain functions to have a much lower sample size; thus, the lowest observed frequency by function is noted by "min" in the titles.

Figure 10: ΔP tests for modifiers by preposition status and genre



These tests show that Time maintains a predictive status for definiteness even when prepositions are removed from the mix (bottom row). In fact, the prediction is generally much higher without the preposition (e.g. +29% for poetry, N=129). However, Loca, which also preferred definiteness, loses its prediction in non-prepositional phrases. Might this be the work of the Masoretes?

There are three reasons to withhold judgment on Loca and definiteness. First, the stipulation of non-prepositional phrases has caused Loca to become significantly underrepresented (as low as N=13 in poetry). This is due to the fact that locative function is intertwined with prepositions in BH.[[98]](#footnote-98) Second, the stipulation has also shifted Loca's prediction from definiteness to genitive in prose and prophecy (+30%, +34%).[[99]](#footnote-99) So it has simply highlighted a different construction, which is irrelevant for the issue of the definite article. Third, there appears to be a convincing functional explanation of why Loca and Time would share this profile, to be explained further.

It is widely recognized that locative adverbials in world languages serve as sources for locating time adverbials.[[100]](#footnote-100) The common definiteness profile of Time and Loca aligns with this hypothesis in a surprising way. There are two dynamics which may explain this. The first concerns the various meanings of the definite article in language. The second concerns the special role that definite articles serve in anchoring time and location references.

Bekins has helpfully summarized the main recognized uses of the definite article. The basic sense of the article is to mark identifiability on a noun.[[101]](#footnote-101) Nouns can be identifiable for a range of reasons, including previous mentions (anaphora) or context (e.g. 'the fork' at 'dinner').[[102]](#footnote-102) One particularly relevant cause is the global use, which marks an item as identifiable due to well-known world or cultural knowledge.[[103]](#footnote-103)

The simple construction [*preposition + article + head*] accounts for 86% of all definite Locas and 25% of all definite Times.[[104]](#footnote-104) The head-word counts for this construction are show below.

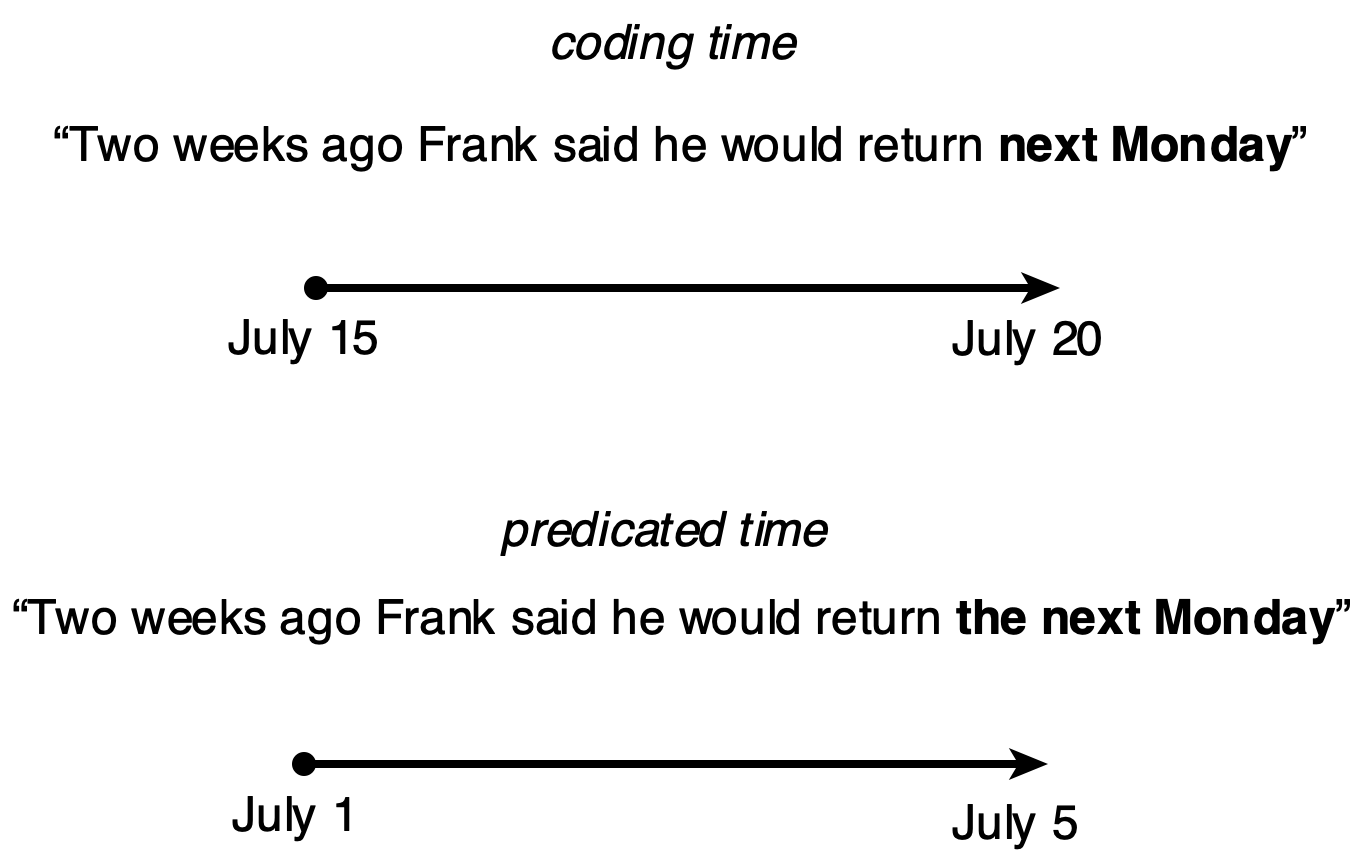
Table 15: Head word frequencies in [*preposition + article + head*] for Time and Loca

|  |  |  |  |
| --- | --- | --- | --- |
| Time head | freq | Loca head | freq |
| **בֹּקֶר** | 100 | **אֶרֶץ** | 64 |
| **עֶרֶב** | 79 | **מִדְבָּר** | 57 |
| **יֹום** | 44 | **שָׂדֶה** | 26 |
| **לַיְלָה** | 23 | **שֶׁמֶשׁ** | 25 |
| **רִאשֹׁון** | 15 | **הַר** | 24 |
| **צָהֳרַיִם** | 12 | **שָׁמַיִם** | 20 |
| **שָׁנָה** | 5 | **דֶּרֶךְ** | 19 |
| **מֹועֵד** | 5 | **עִיר** | 18 |
| **עֵת** | 5 | **בַּיִת** | 18 |
| **חֹדֶשׁ** | 4 | **מִזְבֵּחַ** | 14 |

Both the Time and Loca heads are semantically very general terms which are known either experientially (e.g. natural cycles of the sun) or culturally (e.g. **מוֹעֵד** 'appointed time'). The Time heads align with what Haspelmath calls "canonical times," which are pervasive in world languages.[[105]](#footnote-105) Generic locations are also seen, especially with the noun **שֶׁמֶשׁ** (all Eccl, **תַּחַת הַשֶׁמֶשׁ**). In response to this data, I manually tagged the 838 cases of [*preposition + article + head*] for the major definiteness categories.[[106]](#footnote-106) The tagging confirms that 85% of Time and 54% of Loca represent a global use. Thus, the need to relate time and locations to well-known objects or phenomena appears to provide a motivation for definiteness in Time and Loca. This situation in nominal time words differs from adverb-associated words, which carry innate, deictic reference.[[107]](#footnote-107)

Allen and Hill recognized that the definite article (in languages that have them) can play an important role in anchoring location and time phrases.[[108]](#footnote-108) The deployment of a definite article can cause the referenced position to shift to a different vantage point.[[109]](#footnote-109) They label two cases "coding locus" (shared context), which are null, and "predicated locus" (external context) with the article. They provide the following illuminative example:

Figure 11: Coding versus predicated time markers[[110]](#footnote-110)



In the first sentence, "next Monday" is unmarked; thus, "the temporal location of the participants is used as the controlling point in establishing a referent" (coding time). However "the next Monday" portrays a position anchored to Frank's speech instead of the situation shared by the speaker and recipient (predicated time).[[111]](#footnote-111) Allen and Hill provide a cognitive motivation for this dynamic: zero-marking corresponds with the default here-and-now, whereas definite marking expresses "'there' or 'then'".[[112]](#footnote-112) Consequently, the indication of definiteness in time/location adverbials sits on a similar axis as the marking of verb tense.[[113]](#footnote-113)

Allen and Hill's analysis leaves several unanswered questions as it pertains to the Hebrew data. For instance, there is a different tolerance for coding time (Ø) depending on the preposition used in BH:

Table 16: Preposition collocation frequencies with nominal or Ø markings

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **בְּ** | **עַד** | **לְ** | **מִן** | **כְּ** | **אַחַר** |
| *nom* | 1064 | 186 | 123 | 66 | 39 | 38 |
| *Ø* | 5 | 19 | 3 | 18 | 2 | 0 |

Table 17: Preposition collocation proportions (rounded)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **בְּ** | **עַד** | **לְ** | **מִן** | **כְּ** | **אַחַר** |
| *nom* | 1.0 | 0.91 | 0.98 | 0.79 | 0.95 | 1.0 |
| *Ø* | 0.0 | 0.09 | 0.02 | 0.21 | 0.05 | 0.0 |

**בְּ** occurs with coding time only 0.5% (5 cases) of the time, whereas **עַד** 'until' and **מִן** 'since' occur 9% and 21% with Ø, respectively. Perhaps this because **בְּ** typically marks a position in the past or future relative to the speaker (e.g. **בַּיּוֹם הַהוּא** 'in that day'). **עַד** or **מִן**, on the other hand, can communicate a situation which overlaps with the present, leaving the deictic, coding time available for selection. A representative case in Exod 18 seems to align with Allen and Hill's analysis:

Exodus 18:13-14

|  |  |
| --- | --- |
| **וַיְהִי֙ מִֽמָּחֳרָ֔ת וַיֵּ֥שֶׁב מֹשֶׁ֖ה לִשְׁפֹּ֣ט אֶת־הָעָ֑ם וַיַּעֲמֹ֤ד הָעָם֙ עַל־מֹשֶׁ֔ה מִן־הַבֹּ֖קֶר עַד־הָעָֽרֶב׃**  **וַיַּרְא֙ חֹתֵ֣ן מֹשֶׁ֔ה אֵ֛ת כָּל־אֲשֶׁר־ה֥וּא עֹשֶׂ֖ה לָעָ֑ם וַיֹּ֗אמֶר מָֽה־הַדָּבָ֤ר הַזֶּה֙ אֲשֶׁ֨ר אַתָּ֤ה עֹשֶׂה֙ לָעָ֔ם מַדּ֗וּעַ אַתָּ֤ה יוֹשֵׁב֙ לְבַדֶּ֔ךָ וְכָל־הָעָ֛ם נִצָּ֥ב עָלֶ֖יךָ מִן־בֹּ֥קֶר עַד־עָֽרֶב׃** | "It was the next day when Moses sat to judge the people. And the people stood before Moses from the morning until the evening.  Moses's father-in-law saw all that he was doing for the people. And he said: 'What is this thing you are doing for the people? Why do you sit by yourself while all the people stand before you from morning until evening?" |

When **בֹּקֶר** and **עֶרֶב** are used in the narrative (non-deictic) they appear with **הַ** 'the'; but when the text shifts to Jethro's perspective, the adverbials switch to coding locus (Ø). Similar, but more complicated cases can be found elsewhere in Exodus.[[114]](#footnote-114)

Further work is needed to determine the unique constraints on coding/predicating time locus in Hebrew.[[115]](#footnote-115) But it seems safe to assume that a higher prevalence of predicating locus in BH nominal adverbials would lead to a higher prediction for definiteness. As Allen and Hill note that this dynamic functions likewise in location phrases, the Loca definiteness identified here probably cannot be fully attributed to Tiberian influence.

The role of definite modifiers as referential anchors sheds light on the other attracted modifiers, namely, demonstratives, ordinals, and genitives. Namely, non-quantifying modifiers provide other landmarks to which a given time word might be anchored.[[116]](#footnote-116) A few representative examples will suffice.

Table 18: Selected examples of Time anchored by demonstratives, ordinals, and genitives

| ref | phrase | sentence | translation |
| --- | --- | --- | --- |
| *Gen 7:11* | בַּיֹּ֣ום הַזֶּ֗ה | בַּיֹּ֣ום הַזֶּ֗ה נִבְקְעוּ֙ כָּֽל־מַעְיְנֹת֙ תְּהֹ֣ום רַבָּ֔ה | On this day all the springs of the great deep were rent open. |
| *Joel 3:2* | בַּיָּמִ֣ים הָהֵ֔מָּה | וְגַ֥ם עַל־הָֽעֲבָדִ֖ים וְעַל־הַשְּׁפָחֹ֑ות בַּיָּמִ֣ים הָהֵ֔מָּה אֶשְׁפֹּ֖וךְ אֶת־רוּחִֽי׃ | And indeed in those days I will pour out my spirit upon the servants and handmaids. |
| *Esth 6:1* | בַּלַּ֣יְלָה הַה֔וּא | בַּלַּ֣יְלָה הַה֔וּא נָדְדָ֖ה שְׁנַ֣ת הַמֶּ֑לֶךְ | On that night the king's sleep fled. |
| *Deut 15:12* | בַשָּׁנָה֙ הַשְּׁבִיעִ֔ת | וּבַשָּׁנָה֙ הַשְּׁבִיעִ֔ת תְּשַׁלְּחֶ֥נּוּ חָפְשִׁ֖י מֵעִמָּֽךְ׃ | In the seventh year you shall set him free from your service. |
| *Jer 36:9* | בַּחֹ֣דֶשׁ הַתְּשִׁעִ֔י | וַיְהִ֣י בַשָּׁנָ֣ה הַ֠חֲמִשִׁית לִיהֹויָקִ֨ים בֶּן־יֹאשִׁיָּ֤הוּ מֶֽלֶךְ־יְהוּדָה֙ בַּחֹ֣דֶשׁ הַתְּשִׁעִ֔י | It was on the fifth year of Jehoiakim son of Joshiah, King of Judah, on the ninth month. |
| *Gen 2:17* | בְּיֹ֛ום אֲכָלְךָ֥ מִמֶּ֖נּוּ | כִּ֗י בְּיֹ֛ום אֲכָלְךָ֥ מִמֶּ֖נּוּ מֹ֥ות תָּמֽוּת׃ | For on the day you eat of it, you shall surely die. |
| *Num 35:32* | עַד־מֹ֖ות הַכֹּהֵֽן׃ | וְלֹא־תִקְח֣וּ כֹ֔פֶר לָנ֖וּס אֶל־עִ֣יר מִקְלָטֹ֑ו לָשׁוּב֙ לָשֶׁ֣בֶת בָּאָ֔רֶץ עַד־מֹ֖ות הַכֹּהֵֽן׃ | You shall not take a bribe for anyone who has fled to a city for refuge, that he might return to dwell in the land, until the death of the priest. |
| *Hos 10:9* | מִימֵי֙ הַגִּבְעָ֔ה | מִימֵי֙ הַגִּבְעָ֔ה חָטָ֖אתָ יִשְׂרָאֵ֑ל | Since the days of the hills you Israel have sinned. |
| *Ps 56:10* | בְּיֹ֣ום אֶקְרָ֑א | אָ֥֨ז יָ֘שׁ֤וּבוּ אֹויְבַ֣י אָ֭חֹור בְּיֹ֣ום אֶקְרָ֑א | Then my enemies will turn back on the day I call. |

The last line of inquiry, encompassing the final group of modifiers, is the weak prediction of Time for prepositional modification. The tendency is especially marked given the strong preposition prediction witnessed with Adju, Loca, and Cmpl.

Table 19: Function and preposition collocation frequencies

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| preposition | Adju | Cmpl | Loca | Objc | Subj | Time |
| *False* | 277 | 713 | 126 | 9096 | 15317 | 842 |
| *True* | 4967 | 11877 | 1276 | 4249 | 76 | 1579 |

Table 20: Function and preposition collocation proportions (rounded)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| preposition | Adju | Cmpl | Loca | Objc | Subj | Time |
| *False* | 0.05 | 0.06 | 0.09 | 0.68 | 1.0 | 0.35 |
| *True* | 0.95 | 0.94 | 0.91 | 0.32 | 0.0 | 0.65 |

The lower use of prepositions causes Time to be placed closer to Objc in PCA plot (Figure 6). These phrases are either quantified time spans or demonstrative uses of the definite article, as the examples below illustrate.

Table 21: Selected examples of non-prepositional Time

|  |  |  |  |
| --- | --- | --- | --- |
| ref | text | sentence | translation |
| Gen 3:14 | כָּל־יְמֵ֥י חַיֶּֽיךָ׃ | וְעָפָ֥ר תֹּאכַ֖ל כָּל־יְמֵ֥י חַיֶּֽיךָ׃ | Dust you shall eat all the days of your life. |
| Exod 15:22 | שְׁלֹֽשֶׁת־יָמִ֛ים | וַיֵּלְכ֧וּ שְׁלֹֽשֶׁת־יָמִ֛ים בַּמִּדְבָּ֖ר | They walked for three days in the desert. |
| 1 Sam 6:1 | שִׁבְעָ֥ה חֳדָשִֽׁים׃ | וַיְהִ֧י אֲרֹון־יְהוָ֛ה בִּשְׂדֵ֥ה פְלִשְׁתִּ֖ים שִׁבְעָ֥ה חֳדָשִֽׁים׃ | The ark of YHWH was in a field of the Philistines for seven months. |
| 2 Kgs 24:18 | אַחַ֤ת עֶשְׂרֵה֙ שָׁנָ֔ה | וְאַחַ֤ת עֶשְׂרֵה֙ שָׁנָ֔ה מָלַ֖ךְ בִּירוּשָׁלִָ֑ם | For eleven years he reigned in Jerusalem. |
| Ps 37:26 | כָּל־הַ֭יֹּום | כָּל־הַ֭יֹּום חֹונֵ֣ן | All the time he is gracious. |
| Dan 10:13 | עֶשְׂרִ֣ים וְאֶחָ֣ד יֹ֔ום | וְשַׂ֣ר׀ מַלְכ֣וּת פָּרַ֗ס עֹמֵ֤ד לְנֶגְדִּי֙ עֶשְׂרִ֣ים וְאֶחָ֣ד יֹ֔ום | The prince of the kingdom of Persia stood against me for twenty-one days. |
| Gen 4:14 | הַיֹּ֗ום | הֵן֩ גֵּרַ֨שְׁתָּ אֹתִ֜י הַיֹּ֗ום מֵעַל֙ פְּנֵ֣י הָֽאֲדָמָ֔ה | Take notice, I am driving you out this day from upon the land. |
| 2 Sam 3:39 | הַיֹּ֥ום | וְאָנֹכִ֨י הַיֹּ֥ום רַךְ֙ | I am tender this day. |
| 1 Kgs 12:7 | הַ֠יֹּום | אִם־הַ֠יֹּום תִּֽהְיֶה־עֶ֜בֶד לָעָ֤ם הַזֶּה֙ | If this day you would be a servant to this people... |
| Jer 34:15 | הַיֹּ֗ום | וַתָּשֻׁ֨בוּ אַתֶּ֜ם הַיֹּ֗ום | You turned back this day. |

Such cases represent adverbial accusative times, which are frequently recognized in the grammars.[[117]](#footnote-117) The closer placement of Time towards Objc corroborates this hypothesis. They also back up a parallel claim by Haspelmath that unmarked noun phrases are derived on analogy with the object role.[[118]](#footnote-118) These data thus confirm a broadly two-part division between nominal time adverbials into prepositional and accusative groups.

**Conclusions**

Using a cognitive-statistical method, this article has examined the collocational tendencies of a large sample of phrases marked for adverbial time function (labeled 'Time') as well as adverbial location, other adverbials, complements, objects, and subjects. The article experimented with a data-driven approach to parts-of-speech for Time heads. A broad division between 'adverb' and 'nominal' classes is confirmed, but with a good degree of variability in between. The data suggests gradient categories better capture the behavior of BH words. Finally, the analysis showed how various nominal modifiers such as the definite article or quantifiers are statistically associated with particular functions. Definiteness was found to be particularly associated with Time due to its use for anchoring time references to well-known entities. The possibility of Tiberian interference in definite prepositional phrases is considered and excluded as a factor for Time phrases. Finally, the analysis finds a lower prediction for prepositions than observed with other adjunctive arguments. This tendency reflects the adverbial accusative lineage present in time adverbials. These modifier data reflect the degree to which constructions in language are not neutral movable pieces, but rather interrelated parts of a complex, interconnected network.

**Works Cited**

Allen, Robert L., and Clifford A. Hill. “Contrast between ø and the in Spatial and Temporal Predication.” *Lingua* 48, no. 2–3 (June 1979): 123–46. https://doi.org/10.1016/0024-3841(79)90002-0.

Barr, James. *Biblical Words for Time*. London: SCM Press, 2005.

———. “‘Determination’ and the Definite Article in Biblical Hebrew.” *Journal of Semitic Studies* 34, no. 2 (1989): 307–35.

Bauer, Hans, Pontus Leander, and Paul Kahle. *Historische Grammatik Der Hebräischen Sprache Des Alten Testamentes*. Halle A.S.: M. Niemeyer, 1922.

Bekins, Peter. “Non-Prototypical Uses of the Definite Article in Biblical Hebrew.” *Journal of Semitic Studies* 58, no. 2 (October 1, 2013): 225–40. https://doi.org/10.1093/jss/fgt001.

———. “The Omission of the Definite Article in Biblical Poetry.” San Antonio, 2016. https://www.academia.edu/30145121/The\_Omission\_of\_the\_Definite\_Article\_in\_Biblical\_Poetry.

Brin, Gershon. *The Concept of Time in the Bible and the Dead Sea Scrolls*. Studies on the Texts of the Desert of Judah 39. Leiden: Brill, 2001.

Croft, William. “A Conceptual Framework for Grammatical Categories (or: A Taxonomy of Propositional Acts).” *Journal of Semantics*, no. 7 (1990): 245–79.

———. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford ; New York: Oxford University Press, 2001.

Cuzzolin, Pierluigi, Putzu Ignazio, and Ramat Paolo. “The Indo-European Adverb in Diachronic and Typological Perspective.” *Indogermanische Forschungen* 111 (2006): 1–38.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” *ArXiv:1810.04805 [Cs]*, May 24, 2019. http://arxiv.org/abs/1810.04805.

DeVries, Simon J. *Yesterday, Today and Tomorrow: Time and History in the Old Testament*. USA: Eerdmans, 1975.

Ellis, Nick C. “Language Acquisition as Rational Contingency Learning.” *Applied Linguistics* 27, no. 1 (March 1, 2006): 1–24. https://doi.org/10.1093/applin/ami038.

Ellis, Nick C., Matthew Brook O’Donnell, and Ute Römer. “Usage-Based Language: Investigating the Latent Structures That Underpin Acquisition: Usage-Based Language.” *Language Learning* 63 (March 2013): 25–51.

Fedorenko, Evelina, Idan Asher Blank, Matthew Siegelman, and Zachary Mineroff. “Lack of Selectivity for Syntax Relative to Word Meanings throughout the Language Network.” *Cognition* 203 (October 2020): 104348. https://doi.org/10.1016/j.cognition.2020.104348.

Fillmore, Charles. “Mini-Grammars of Some Time-When Expressions in English.” In *Complex Sentences in Grammar and Discourse: Essays in Honor of Sandra A. Thompson*, edited by Joan L. Bybee, Sandra A. Thompson, and Michael Noonan. Amsterdam: Benjamins, 2002.

Firth, J.R. “A Synopsis of Linguistic Theory, 1930–1995.” In *Studies in Linguistic Analysis*, 1–32. Oxford: Basil Blackwell, 1962.

Forbes, A. Dean. “Distributionally-Inferred Word and Form Classes in the Hebrew Lexicon: Known by the Company They Keep.” In *Foundations for Syriac Lexicography II*, edited by Peter J. Williams, 1–34. Piscataway, NJ: Gorgias Press, 2009.

———. “Squishes, Clines, and Fuzzy Signs: Mixed and Gradient Categories in the Biblical Hebrew Lexicon.” In *Foundations for Syriac Lexicography: Colloquia of the International Syriac Language Project*, edited by Beryl Turner, A. Dean Forbes, David G. K. Taylor, J. W. Dyk, and W. Th van Peursen, 105–39. Perspectives on Syriac Linguistics 1. Piscataway, NJ: Gorgias Press, 2005.

Geeraerts, Dirk. “The Doctor and the Semantician.” In *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, edited by Dirk Geeraerts and John R. Taylor, 63–78. Cognitive Linguistics Research 46. Berlin: De Gruyter Mouton, 2010.

———. *Theories of Lexical Semantics*. Oxford ; New York: Oxford University Press, 2010.

Geeraerts, Dirk, and Hubert Cuyckens. “Introducing Cognitive Linguistics.” In *The Oxford Handbook of Cognitive Linguistics*, edited by Dirk Geeraerts and Hubert Cuyckens. Oxford: Oxford University Press, 2012.

Gesenius, Wilhelm, Emiel Friedrich Kautsch, and A.E. Cowley. *Gesenius’ Hebrew Grammar*. Second. Oxford: Clarendon Press, 1909.

Goldberg, Adele E., Devin M. Casenhiser, and Nitya Sethuraman. “Learning Argument Structure Generalizations.” *Cognitive Linguistics* 15, no. 3 (2004): 289–316.

Gries, Stefan Th. “Dispersions and Adjusted Frequencies in Corpora.” *International Journal of Corpus Linguistics* 13, no. 4 (2008): 403–37.

Hardy, Humphrey, and Samuel Boyd. “Hebrew Adverbialization, Aramaic Language Contact, and Mpny ʾšr in Exodus 19:18.” In *Semitic Languages in Contact*, edited by Aaron Butts. Brill, 2015. https://doi.org/10.1163/9789004300156.

Harkness, Janet. “Time Adverbials in English and Reference Time.” In *Essays on Tensing in English*, edited by Alfred Schopf, 71–110. Linguistische Arbeiten 185, 228. Tübingen: Max Niemeyer Verlag, 1987.

Haspelmath, Martin. *From Space to Time: Temporal Adverbials in the World’s Languages*. LINCOM Studies in Theoretical Linguistics 2. Münchn: Lincom Europa, 1997.

———. “Pre-Established Categories Don’t Exist: Consequences for Language Description and Typology.” *Linguistic Typology* 11 (2007): 119–32.

Joüon, Paul, and T. Muraoka. *A Grammar of Biblical Hebrew*. Subsidia Biblica, 14/1-14/2. Roma: Editrice Pontificio Istituto Biblio, 1996.

Kingham, Cody. *CambridgeSemiticsLab/BH\_time\_collocations: Time Collocations with Article Data* (version 4.1). Zenodo, 2020. https://doi.org/10.5281/ZENODO.3931953.

Kingham, Cody, and Wido Van Peursen. “The ETCBC Database of the Hebrew Bible.” *Journal for Semitics* 27, no. 1 (2018): 1–13. https://doi.org/10.25159/1013-8471/2974.

Klein, Wolfgang. *Time in Language*. Germanic Linguistics. London: Routledge, 1994.

Köhler, Ludwig, and Walter Baumgartner. *Supplementum Ad Lexicon in Veteris Testamenti Libros*. Leiden: Brill, 1958.

Kuhn, Thomas S. “The Natural and the Human Sciences.” In *The Road Since Structure: Philosophical Essays, 1970–1993, with an Autobiographical Interview*, edited by James Conant and John Haugeland, 216–23. Chicago: University of Chicago Press, 2000.

Lakoff, George. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago: University of Chicago Press, 1987.

Levshina, Natalia. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins, 2015.

Ley, Julius. “Über Den Gebrauch Des Artikels in Der Rhythmischen Poesie Der Hebräer.” *Neue Jahrbücher Für Philologie Und Paedagogik* 144, no. 2 (1891): 341–451.

Locatell, Christian. “Grammatical Polysemy in the Hebrew Bible: A Cognitive Linguistic Approach to כי.” PhD Dissertation, University of Stellenbosch, 2017.

Lyavdansky, Alexey. “Temporal Deictic Adverbs as Discourse Markers in Hebrew, Aramaic and Akkadian.” *Journal of Language Relationship* 3 (2010): 22–42.

Miller-Naudé, Cynthia L., and Jacobus A. Naudé. “A Re-Examination of Grammatical Categorization in Biblical Hebrew.” In *From Ancient Manuscripts to Modern Dictionaries: Select Studies in Aramaic, Hebrew and Greek*, edited by Society of Biblical Literature, Tarsee Li, and Keith D. Dyer, 273–308. Perspectives on Linguistics and Ancient Languages 9. Piscataway, NJ: Gorgias Press, 2017.

Niccacci, Alviero. “Temporal Clause: Biblical Hebrew.” Edited by Geoffrey Khan. *Encyclopedia of Hebrew Language and Linguistics*. Koninklijke Brill NV, 2013. https://doi.org/10.1163/2212-4241\_ehll\_EHLL\_COM\_00000598.

Oosting, Reinoud. “Computer-Assisted Analysis of Old Testament Texts: The Contribution of the WIVU to Old Testament Scholarship.” In *The Present State of Old Testament Studies in the Low Countries: A Collection of Old Testament Studies Published on the Occasion of the Seventy-Fifth Anniversary of the Oudtestamentisch Werkgezelschap*, edited by Klaas Spronk, 69:192–209. Oudtestamentische Studiën. Leiden: Brill, 2016.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. “Scikit-Learn: Machine Learning in Python.” *Journal of Machine Learning Research* 12 (2011): 2825–2830.

Perek, Florent, and Adele E. Goldberg. “Linguistic Generalization on the Basis of Function and Constraints on the Basis of Statistical Preemption.” *Cognition* 168 (2017): 276–93. https://doi.org/10.1016/j.cognition.2017.06.019.

Piantadosi, Steven T. “Zipf’s Word Frequency Law in Natural Language: A Critical Review and Future Directions.” *Psychonomic Bulletin & Review* 21, no. 5 (October 2014): 1112–30. https://doi.org/10.3758/s13423-014-0585-6.

Ramat, Paolo. “Adverbial Grammaticalization.” In *The Oxford Handbook of Grammaticalization*, edited by Bernd Heine and Heiko Narrog. Oxford University Press, 2011. https://doi.org/10.1093/oxfordhb/9780199586783.013.0040.

Ravid, Dorit, and Yitzhak Shlesinger. “Modern Hebrew Adverbials: Between Syntactic Class and Lexical Category.” In *Between Grammar and Lexicon*, edited by Ellen Contini-Morava and Y. Tobin, 333–51. Amsterdam Studies in the Theory and History of Linguistic Science, v. 183. Amsterdam: John Benjamins, 2000.

Rokhlin, Vladimir, Arthur Szlam, and Mark Tygert. “A Randomized Algorithm for Principal Component Analysis.” *SIAM Journal on Matrix Analysis and Applications* 31, no. 3 (January 2010): 1100–1124. https://doi.org/10.1137/080736417.

Roorda, Dirk. “The Hebrew Bible as Data: Laboratory - Sharing - Experiences.” In *Clarin in the Low Countries*, edited by J. Odijk and A. Van Hessen. London: Ubiquity Press, 2017.

Roorda, Dirk, Constantijn Sikkel, Wido Van Peursen, Martijn Naaijer, Christiaan Erwich, and Cody Kingham. “Biblia Hebraica Stuttgartensia (Amstelodamensis) Documentation.” Documentation. ETCBC Github, April 10, 2019. https://etcbc.github.io/bhsa/.

Sáenz-Badillos, Angel. *A History of the Hebrew Language*. Reprinted. Cambridge: Cambridge University Press, 1997.

Scholz, Barbara C., Francis Jeffry Pelletier, and Geoffrey K. Pullum. “Philosophy of Linguistics.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2011. https://plato.stanford.edu/archives/win2016/entries/linguistics/.

Schütze, Carson T. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Classics in Linguistics 2. Berlin: Language Science Press, 2016.

Sloutsky, Vladimir M., Wei (Sophia) Deng, Anna V. Fisher, and Heidi Kloos. “Conceptual Influences on Induction: A Case for a Late Onset.” *Cognitive Psychology* 82 (November 2015): 1–31. https://doi.org/10.1016/j.cogpsych.2015.08.005.

Stefanowitsch, Anatol. *Corpus Linguistics: A Guide to the Methodology*. Textbooks in Language Sciences 7. Berlin: Language Science Press, 2020.

———. “Empirical Cognitive Semantics: Some Thoughts.” In *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, edited by Dylan Glynn and Kertin Fischer, 355–80. Cognitive Linguistics Research 46. Berlin: De Gruyter Mouton, 2010.

Stefanowitsch, Anatol, and Stefan Th. Gries. “Collostructions: Investigating the Interaction of Words and Constructions.” *International Journal of Corpus Linguistics* 8, no. 2 (2003): 209–43.

ttnphns (https://stats.stackexchange.com/users/3277/ttnphns). “Loadings vs Eigenvectors in PCA: When to Use One or Another?” Stack Exchange, March 29, 2015. https://stats.stackexchange.com/q/143949.

Van der Merwe, C. H. J., J. A. Naudé, and Jan Kroeze. *A Biblical Hebrew Reference Grammar*. Second edition. New York: Bloomsbury T&T Clark, 2017.

Van der Merwe, Christo H.J. “Reconsidering Biblical Hebrew Temporal Expressions.” *ZAH* 10, no. 1 (1997): 42–62.

Waltke, Bruce, and M. O’Connor. *An Introduction to Biblical Hebrew Syntax*. Winona Lake: Eisenbrauns, 1990.

Waskom, Michael, Olga Botvinnik, Joel Ostblom, Maoz Gelbart, Saulius Lukauskas, Paul Hobson, David C Gemperline, et al. *Mwaskom/Seaborn: V0.10.1 (April 2020)* (version v0.10.1). Zenodo, 2020. https://doi.org/10.5281/ZENODO.3767070.

Xiao, R. “Collocation.” In *The Cambridge Handbook of English Corpus Linguistics*, edited by D. Biber and R. Reppen, 106–24. Cambridge: Cambridge University, 2015.

1. See, though, the important work of Van der Merwe (1997). There have also been numerous studies dedicated to individual lexemes or to temporal clauses (כִּי 'for/when', עַתָּה 'now'). There have been two large exegetical studies on time phrases (Brin and DeVries). This study focuses instead on time adverbials as a broader linguistic phenomenon. Christo H.J. Van der Merwe, “Reconsidering Biblical Hebrew Temporal Expressions,” *ZAH* 10, no. 1 (1997): 42–62; James Barr, *Biblical Words for Time* (London: SCM Press, 2005); Alviero Niccacci, “Temporal Clause: Biblical Hebrew,” ed. Geoffrey Khan, *Encyclopedia of Hebrew Language and Linguistics* (Koninklijke Brill NV, 2013), https://doi.org/10.1163/2212-4241\_ehll\_EHLL\_COM\_00000598; Alexey Lyavdansky, “Temporal Deictic Adverbs as Discourse Markers in Hebrew, Aramaic and Akkadian,” *Journal of Language Relationship* 3 (2010): 22–42; Christian Locatell, “Grammatical Polysemy in the Hebrew Bible: A Cognitive Linguistic Approach to כי” (PhD Dissertation, Stellenbosch, South Africa, University of Stellenbosch, 2017); Gershon Brin, *The Concept of Time in the Bible and the Dead Sea Scrolls*, Studies on the Texts of the Desert of Judah 39 (Leiden: Brill, 2001); Simon J. DeVries, *Yesterday, Today and Tomorrow: Time and History in the Old Testament* (USA: Eerdmans, 1975). [↑](#footnote-ref-1)
2. C. H. J. Van der Merwe, J. A. Naudé, and Jan Kroeze, *A Biblical Hebrew Reference Grammar*, Second edition (New York: Bloomsbury T&T Clark, 2017), §33.1; Wilhelm Gesenius, Emiel Friedrich Kautsch, and A.E. Cowley, *Gesenius’ Hebrew Grammar*, Second (Oxford: Clarendon Press, 1909), §118a; Bruce Waltke and M. O’Connor, *An Introduction to Biblical Hebrew Syntax* (Winona Lake: Eisenbrauns, 1990), §10.2a. [↑](#footnote-ref-2)
3. Van der Merwe asked similar questions in 1997: "Is there any difference in the syntax of non-temporal adjuncts and temporal adjuncts? Which BH constructions can function as temporal adjuncts? Apart from the above-mentioned semantic classes of temporal position, duration and frequency, are there other semantic classes or subclasses to be identified among temporal adjuncts?" Van der Merwe, “Reconsidering Temporal Expressions,” 49. [↑](#footnote-ref-3)
4. E.g. Cynthia L. Miller-Naudé and Jacobus A. Naudé, “A Re-Examination of Grammatical Categorization in Biblical Hebrew,” in *From Ancient Manuscripts to Modern Dictionaries: Select Studies in Aramaic, Hebrew and Greek*, ed. Society of Biblical Literature, Tarsee Li, and Keith D. Dyer, Perspectives on Linguistics and Ancient Languages 9 (Piscataway, NJ: Gorgias Press, 2017), 273–308; A. Dean Forbes, “Distributionally-Inferred Word and Form Classes in the Hebrew Lexicon: Known by the Company They Keep,” in *Foundations for Syriac Lexicography II*, ed. Peter J. Williams (Piscataway, NJ: Gorgias Press, 2009), 1–34. [↑](#footnote-ref-4)
5. Barbara C. Scholz, Francis Jeffry Pelletier, and Geoffrey K. Pullum, “Philosophy of Linguistics,” in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, 2011, https://plato.stanford.edu/archives/win2016/entries/linguistics/; Dirk Geeraerts, *Theories of Lexical Semantics* (Oxford ; New York: Oxford University Press, 2010). [↑](#footnote-ref-5)
6. Evelina Fedorenko et al., “Lack of Selectivity for Syntax Relative to Word Meanings throughout the Language Network,” *Cognition* 203 (October 2020): 104348, https://doi.org/10.1016/j.cognition.2020.104348; Jacob Devlin et al., “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding,” *ArXiv:1810.04805 [Cs]*, May 24, 2019, http://arxiv.org/abs/1810.04805; Florent Perek and Adele E. Goldberg, “Linguistic Generalization on the Basis of Function and Constraints on the Basis of Statistical Preemption,” *Cognition* 168 (2017): 276–93, https://doi.org/10.1016/j.cognition.2017.06.019. [↑](#footnote-ref-6)
7. Martin Haspelmath, “Pre-Established Categories Don’t Exist: Consequences for Language Description and Typology,” *Linguistic Typology* 11 (2007): 119–32. [↑](#footnote-ref-7)
8. Van der Merwe, Naudé, and Kroeze, *BHRG*, 380–81; Waltke and O’Connor, *IBHS*, 39.3.1a. [↑](#footnote-ref-8)
9. Dirk Geeraerts and Hubert Cuyckens, “Introducing Cognitive Linguistics,” in *The Oxford Handbook of Cognitive Linguistics*, ed. Dirk Geeraerts and Hubert Cuyckens (Oxford: Oxford University Press, 2012), §2. [↑](#footnote-ref-9)
10. I agree with Ellis et al.: "We take the Saussurian (1916) view that the units of language are constructions—form-meaning mappings, conventionalized in the speech community, and entrenched as language knowledge in the learner’s mind." Nick C. Ellis, Matthew Brook O’Donnell, and Ute Römer, “Usage-Based Language: Investigating the Latent Structures That Underpin Acquisition: Usage-Based Language,” *Language Learning* 63 (March 2013): 25–51. For experiential categories see George Lakoff, *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind* (Chicago: University of Chicago Press, 1987). [↑](#footnote-ref-10)
11. Lakoff, *Women, Fire, and Dangerous Things*, 41. [↑](#footnote-ref-11)
12. Steven T. Piantadosi, “Zipf’s Word Frequency Law in Natural Language: A Critical Review and Future Directions,” *Psychonomic Bulletin & Review* 21, no. 5 (October 2014): 1112–30, https://doi.org/10.3758/s13423-014-0585-6; Ellis, O’Donnell, and Römer, “Usage-Based Language,” 31–33. [↑](#footnote-ref-12)
13. On similarity-based categorization see Vladimir M. Sloutsky et al., “Conceptual Influences on Induction: A Case for a Late Onset,” *Cognitive Psychology* 82 (November 2015): 1–31, https://doi.org/10.1016/j.cogpsych.2015.08.005. [↑](#footnote-ref-13)
14. Anatol Stefanowitsch, “Empirical Cognitive Semantics: Some Thoughts,” in *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, ed. Dylan Glynn and Kertin Fischer, Cognitive Linguistics Research 46 (Berlin: De Gruyter Mouton, 2010), 355–80. [↑](#footnote-ref-14)
15. Adele E. Goldberg, Devin M. Casenhiser, and Nitya Sethuraman, “Learning Argument Structure Generalizations,” *Cognitive Linguistics* 15, no. 3 (2004): 289–316. [↑](#footnote-ref-15)
16. Indeed, as noted above, these mental associations are learned over relationships observed in the world. Language thus represents a re-externalization of semantic input. [↑](#footnote-ref-16)
17. Famously summarized by Firth as "knowing a word by the company it keeps." J.R. Firth, “A Synopsis of Linguistic Theory, 1930–1995,” in *Studies in Linguistic Analysis* (Oxford: Basil Blackwell, 1962), 1–32; R. Xiao, “Collocation,” in *The Cambridge Handbook of English Corpus Linguistics*, ed. D. Biber and R. Reppen (Cambridge: Cambridge University, 2015), 106–24. [↑](#footnote-ref-17)
18. Based on an ongoing analysis by this author: https://github.com/codykingham/noun\_semantics/blob/master/analysis.ipynb [↑](#footnote-ref-18)
19. Anatol Stefanowitsch and Stefan Th. Gries, “Collostructions: Investigating the Interaction of Words and Constructions,” *International Journal of Corpus Linguistics* 8, no. 2 (2003): 209–43. [↑](#footnote-ref-19)
20. See, for instance, Devlin et al., “BERT.” [↑](#footnote-ref-20)
21. See Carson T. Schütze, *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*, Classics in Linguistics 2 (Berlin: Language Science Press, 2016). [↑](#footnote-ref-21)
22. See Dirk Geeraerts, “The Doctor and the Semantician,” in *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, ed. Dirk Geeraerts and John R. Taylor, Cognitive Linguistics Research 46 (Berlin: De Gruyter Mouton, 2010), 63–78. [↑](#footnote-ref-22)
23. A. Dean Forbes, “Squishes, Clines, and Fuzzy Signs: Mixed and Gradient Categories in the Biblical Hebrew Lexicon,” in *Foundations for Syriac Lexicography: Colloquia of the International Syriac Language Project*, ed. Beryl Turner et al., Perspectives on Syriac Linguistics 1 (Piscataway, NJ: Gorgias Press, 2005), 105–39; Forbes, “Distributionally-Inferred.” [↑](#footnote-ref-23)
24. Thomas S. Kuhn, “The Natural and the Human Sciences,” in *The Road Since Structure: Philosophical Essays, 1970–1993, with an Autobiographical Interview*, ed. James Conant and John Haugeland (Chicago: University of Chicago Press, 2000), 216–23. [↑](#footnote-ref-24)
25. This is the process known as 'operationalization' and is the "secret" to the success of the natural sciences. Stefanowitsch, “Empirical Cognitive Semantics,” 358–61. [↑](#footnote-ref-25)
26. Dirk Roorda et al., “Biblia Hebraica Stuttgartensia (Amstelodamensis) Documentation,” Documentation, ETCBC Github, April 10, 2019, https://etcbc.github.io/bhsa/. [↑](#footnote-ref-26)
27. Dirk Roorda, “The Hebrew Bible as Data: Laboratory - Sharing - Experiences,” in *Clarin in the Low Countries*, ed. J. Odijk and A. Van Hessen (London: Ubiquity Press, 2017). [↑](#footnote-ref-27)
28. Reinoud Oosting, “Computer-Assisted Analysis of Old Testament Texts: The Contribution of the WIVU to Old Testament Scholarship,” in *The Present State of Old Testament Studies in the Low Countries: A Collection of Old Testament Studies Published on the Occasion of the Seventy-Fifth Anniversary of the Oudtestamentisch Werkgezelschap*, ed. Klaas Spronk, vol. 69, Oudtestamentische Studiën (Leiden: Brill, 2016), 192–209. [↑](#footnote-ref-28)
29. This is done using a custom Python parser which works within the existing phrase segmentation of the BHSA. In some cases, the phrase segmentations are modified where they are deficient. The parser can be viewed in the project repository at https://github.com/CambridgeSemiticsLab/BH\_time\_collocations/blob/master/data/cxs/phrase\_grammar.py [↑](#footnote-ref-29)
30. Anatol Stefanowitsch, *Corpus Linguistics: A Guide to the Methodology*, Textbooks in Language Sciences 7 (Berlin: Language Science Press, 2020), 111–16. [↑](#footnote-ref-30)
31. Stefanowitsch, 113. [↑](#footnote-ref-31)
32. The whole project is archived at the following reference. For this article, data is stored in the various directories under the heading 'advb\_article'. The datasets and production script can be found at data/advb\_article (see function\_data.csv and generate\_function\_data.ipynb). The parsers can be found under /tools/cx\_analysis and data/cxs. The principle analysis code is under results/notebooks/advb\_article. Figures are under results/figures. results/spreadsheets/advb\_article includes the PCA values for the part of speech analysis. The whole directory requires a virtual python environment with /tools appended to the Python namespace. requirements.txt provides the necessary prerequisites. Cody Kingham, *CambridgeSemiticsLab/BH\_time\_collocations: Time Collocations with Article Data*, version 4.1 (Zenodo, 2020), https://doi.org/10.5281/ZENODO.3931953. [↑](#footnote-ref-32)
33. Cody Kingham and Wido Van Peursen, “The ETCBC Database of the Hebrew Bible,” *Journal for Semitics* 27, no. 1 (2018): 1–13, https://doi.org/10.25159/1013-8471/2974. [↑](#footnote-ref-33)
34. I used Croft's notion of semantic headship as a general guide. Objects of prepositions are the semantic heads of prepositional phrases. Quantifiers including כֹל 'all' are excluded as heads if they modify a word. William Croft, *Radical Construction Grammar: Syntactic Theory in Typological Perspective* (Oxford ; New York: Oxford University Press, 2001), 257–59. [↑](#footnote-ref-34)
35. A measure called deviation of proportions (Gries 2008) was used to compare the sampled distribution by book against the whole dataset. The most affected books are Deut (3% total deviation), Ps (3%), 2 Chr (2%), Isa (2%), and Job (2%). Stefan Th. Gries, “Dispersions and Adjusted Frequencies in Corpora,” *International Journal of Corpus Linguistics* 13, no. 4 (2008): 415–19. [↑](#footnote-ref-35)
36. Throughout the study I will refer to ratios out of 1, which can be converted to percentages by multiplying them by 100. This is standard practice in statistical reporting. Natalia Levshina, *How to Do Linguistics with R: Data Exploration and Statistical Analysis* (Amsterdam: John Benjamins, 2015), 70–71. [↑](#footnote-ref-36)
37. This function is more heavily affected by the selection criteria than others. This is due to the exclusion of prepositional phrases that contain only a preposition + suffix. Technically the suffix serves as the head of these phrases, but for now I must exclude these cases since suffixes are not modeled as words in the database. [↑](#footnote-ref-37)
38. Stefanowitsch, “Empirical Cognitive Semantics,” 358–61. [↑](#footnote-ref-38)
39. E.g. Gesenius, Kautsch, and Cowley, *Hebrew Grammar*, §100, 118; Paul Joüon and T. Muraoka, *A Grammar of Biblical Hebrew*, Subsidia Biblica, 14/1-14/2 (Roma: Editrice Pontificio Istituto Biblio, 1996), §102a-c, 126i, 133; Waltke and O’Connor, *IBHS*, §10.2.2c, 11.2, 39.3.1h. [↑](#footnote-ref-39)
40. Van der Merwe, Naudé, and Kroeze, *BHRG*, 380–81; Waltke and O’Connor, *IBHS*, §39.3.1a. [↑](#footnote-ref-40)
41. A handful of terms possess adverb suffixes such as ָם as found on **יוֹמָם** 'by day'. Van der Merwe, Naudé, and Kroeze, *BHRG*, 380. [↑](#footnote-ref-41)
42. Pierluigi Cuzzolin, Putzu Ignazio, and Ramat Paolo, “The Indo-European Adverb in Diachronic and Typological Perspective,” *Indogermanische Forschungen* 111 (2006): 1–38. [↑](#footnote-ref-42)
43. Croft, *Radical Construction Grammar*, 97; Dorit Ravid and Yitzhak Shlesinger, “Modern Hebrew Adverbials: Between Syntactic Class and Lexical Category,” in *Between Grammar and Lexicon*, ed. Ellen Contini-Morava and Y. Tobin, Amsterdam Studies in the Theory and History of Linguistic Science, v. 183 (International Cognitive Linguistics Conference, Amsterdam: John Benjamins, 2000), 335–37. [↑](#footnote-ref-43)
44. Paolo Ramat, “Adverbial Grammaticalization,” in *The Oxford Handbook of Grammaticalization*, ed. Bernd Heine and Heiko Narrog (Oxford University Press, 2011), https://doi.org/10.1093/oxfordhb/9780199586783.013.0040. [↑](#footnote-ref-44)
45. Ravid and Shlesinger, “Modern Hebrew Adverbials,” 339–44; Van der Merwe, Naudé, and Kroeze, *BHRG*, 380–81. [↑](#footnote-ref-45)
46. Croft notes Nootkan, Salishan, Iroquoian, Philippine, and Polynesian languages as oft-cited examples. Croft, *Radical Construction Grammar*, 65. [↑](#footnote-ref-46)
47. Croft, 63–107. [↑](#footnote-ref-47)
48. Croft, 86–87. [↑](#footnote-ref-48)
49. William Croft, “A Conceptual Framework for Grammatical Categories (or: A Taxonomy of Propositional Acts),” *Journal of Semantics*, no. 7 (1990): 245–79. [↑](#footnote-ref-49)
50. Copied from Croft, 248. [↑](#footnote-ref-50)
51. Croft, *Radical Construction Grammar*, 98–99. [↑](#footnote-ref-51)
52. Croft, “A Conceptual Framework,” 256–60. [↑](#footnote-ref-52)
53. i.e. relative to speech time [↑](#footnote-ref-53)
54. Martin Haspelmath, *From Space to Time: Temporal Adverbials in the World’s Languages*, LINCOM Studies in Theoretical Linguistics 2 (Münchn: Lincom Europa, 1997), 43–44; Wolfgang Klein, *Time in Language*, Germanic Linguistics (London: Routledge, 1994), 150–58. [↑](#footnote-ref-54)
55. Croft, “A Conceptual Framework,” 271. [↑](#footnote-ref-55)
56. Croft, 253. [↑](#footnote-ref-56)
57. Croft, 279. [↑](#footnote-ref-57)
58. This includes articles only reflected in the vocalized text of BHS. [↑](#footnote-ref-58)
59. In this case the modifier term is the *nomen rectum*, or trailing term in a construct chain, rather than the head which is morphologically marked as construct. For this reason the term 'genitive' is used. The BHSA dataset includes constructs indicated via the accents or context. [↑](#footnote-ref-59)
60. The attributive adjunct construction involving the definite article, i.e. הַ + word + הַ + adjunct, is also tagged separately. But since that construction itself involves definite modification, it is excluded from this study. The simple adjunct construction of word + adjunct is not yet tagged due to existing limitations in the dataset. In the BHSA, these relations are tagged as so-called 'subphrase relations'; but subphrase relations suffer from a number of problems and inadequacies making them unreliable. In the case of time adverbials, it seems (based on previous analyses) that adjectival elements play a lesser role in the semantics of time. Nevertheless, in the future it would be better to include these modifiers in the dataset. [↑](#footnote-ref-60)
61. An alternative normalization would be a measure of statistical significance as used elsewhere in this study. But in this case, a ratio makes more sense because we want to compare word tendencies to one another rather than find particular distinctives of each individual word. [↑](#footnote-ref-61)
62. Principle component analysis is a dimensionality reduction method which seeks to maximize the variation between the variable vectors (i.e. lists of observed ratios), and thus separate samples along an axis of greatest variation (the first principle component) and second greatest variation (the second). The first principle component can be used as an x-axis value and the second as the y-axis value, yielding a graph of samples which are placed nearer or farther to one another based on their features. Vladimir Rokhlin, Arthur Szlam, and Mark Tygert, “A Randomized Algorithm for Principal Component Analysis,” *SIAM Journal on Matrix Analysis and Applications* 31, no. 3 (January 2010): 1100–1124, https://doi.org/10.1137/080736417; F. Pedregosa et al., “Scikit-Learn: Machine Learning in Python,” *Journal of Machine Learning Research* 12 (2011): 2825–2830. [↑](#footnote-ref-62)
63. The lexicon parts of speech values come from the BHSA's 'sp' feature, which is derived from KBL2. The abbreviated values are substantive, interrogative, adjective, and adverb, respectively. Ludwig Köhler and Walter Baumgartner, *Supplementum Ad Lexicon in Veteris Testamenti Libros* (Leiden: Brill, 1958). [↑](#footnote-ref-63)
64. These are the loading scores, which are the following the discussion of ttnphns (https://stats.stackexchange.com/users/3277/ttnphns), “Loadings vs Eigenvectors in PCA: When to Use One or Another?,” Stack Exchange, March 29, 2015, https://stats.stackexchange.com/q/143949. [↑](#footnote-ref-64)
65. The loading score includes direction from the eigen vector and magnitude from the eigen value. The Ø modifier has a PC1 (x-axis) of -4.8 and PC2 (y-axis) of 3.9. The line is drawn from origin to this point. [↑](#footnote-ref-65)
66. E.g. Gen 7:11; Deut 3:14; Josh 20:6; 1 Sam 7:6; 2 Sam 3:38; 2 Kgs 20:1; Isa 28:5; Ezek 39:11; Zech 13:4; 2 Chr 35:16. [↑](#footnote-ref-66)
67. E.g. Isa 47:12, 15; Judg 16:30, 2 Sam 18:18. [↑](#footnote-ref-67)
68. In Modern Hebrew the distinction between adjectives and adverbs is blurry. A similar dynamic appears in BH. Ravid and Shlesinger, “Modern Hebrew Adverbials,” 346. [↑](#footnote-ref-68)
69. Compared to words with physical manifestations like יוֹם 'day' or עֶרֶב 'evening'. [↑](#footnote-ref-69)
70. Jer 20:3, 2 Kgs 8:15, 1 Sam 31:8, 1 Sam 18:10, 1 Sam 11:11, Judg 21:4, Judg 9:42, 1 Chr 10:8, Num 17:23, Exod 32:30, Exod 18:13. [↑](#footnote-ref-70)
71. See the subsequent section regarding the status of vocalized articles in the Tiberian tradition. [↑](#footnote-ref-71)
72. Thanks to Chip Hardy for pointing out these possibilities in relation to another similar case, לְעוֹלָם. [↑](#footnote-ref-72)
73. N.B. that this table contains a higher marginal total than the previous since modifiers can co-occur. [↑](#footnote-ref-73)
74. The closest would be בֹּקֶר with 18/119 (15%) null modified uses, a substantial difference from the 35% of לַיְלָה. יוֹם has 0.08% Ø; עֶרֶב has 3.5% Ø. [↑](#footnote-ref-74)
75. Thanks to Chip Hardy for pointing this out. Hans Bauer, Pontus Leander, and Paul Kahle, *Historische Grammatik Der Hebräischen Sprache Des Alten Testamentes* (Halle A.S.: M. Niemeyer, 1922), 528. [↑](#footnote-ref-75)
76. Another case is נֵצַח 'everlasting' with 96% null modification (27/28) and 1 nominal construal in Isa 34:10: **לְנֵ֣צַח נְצָחִ֔ים אֵ֥ין עֹבֵ֖ר בָּֽהּ** [↑](#footnote-ref-76)
77. Angel Sáenz-Badillos, *A History of the Hebrew Language*, Reprinted (Cambridge: Cambridge University Press, 1997), 200. [↑](#footnote-ref-77)
78. Thanks again to Chip Hardy for this idea. [↑](#footnote-ref-78)
79. Josh 24:2; Isa 42:14; Isa 57:11; Isa 63:19; Isa 64:3; Jer 2:20; Prov 8:23. [↑](#footnote-ref-79)
80. E.g. Miller-Naudé and Naudé make this argument with respect to הטּוֹב 'the good'. Miller-Naudé and Naudé, “A Re-Examination,” 288–303. [↑](#footnote-ref-80)
81. Nick C. Ellis, “Language Acquisition as Rational Contingency Learning,” *Applied Linguistics* 27, no. 1 (March 1, 2006): 10–12, https://doi.org/10.1093/applin/ami038. [↑](#footnote-ref-81)
82. Contingency-based tests uses 4 variables. Given a construction X, a co-occurring construction Y, and a frequency calculator N, the 4 variables are:

    A = N(X with Y)

    B = N(X) - A

    C = N(Y) - A

    D = N(¬X|¬Y) - (A+B+C), where ¬ means 'not'.

    Given these values, ΔP can be calculated as . Levshina, *Linguistics with R*, 223, 234. [↑](#footnote-ref-82)
83. Ellis, “Language Acquisition,” 11. [↑](#footnote-ref-83)
84. That is, the probability of a response given the cue minus the probability of the response without the cue. Ellis, 11. [↑](#footnote-ref-84)
85. A True/False value is stored on every instance in the sample for whether a word contains a given nominal modifier. [↑](#footnote-ref-85)
86. A lexeme that occurs twice, for instance, might have 1 nominal modifier and 1 Ø, yielding a ratio of 0.5 (50%). [↑](#footnote-ref-86)
87. This is also a vindication of using a statistical method rather than a lexicon label since certain conceptual 'nouns' (e.g. proper names) may not have proto-typical noun behavior. [↑](#footnote-ref-87)
88. The genre labels were made manually during the ETCBC's recent syntactic variation project. They are intended to be coarse-grained. Books receive an overall label (e.g. 'prose' for Gen); that label is then overridden in specific stretches (i.e. 'prose' becomes 'list' in Gen 5, and 'poetry' in Gen 49). See the discussion in <https://github.com/ETCBC/genre_synvar/issues/1> [↑](#footnote-ref-88)
89. Note that the marginal frequency for the whole table is greater than the sample size since modifiers may co-occur in the same phrase. [↑](#footnote-ref-89)
90. Seaborn was used. Michael Waskom et al., *Mwaskom/Seaborn: V0.10.1 (April 2020)*, version v0.10.1 (Zenodo, 2020), https://doi.org/10.5281/ZENODO.3767070. [↑](#footnote-ref-90)
91. Typically the x-axis would contain the independent variables. But in this case, I found the values to be more readable from y to x-axis, indicating "y cues x". [↑](#footnote-ref-91)
92. This measure is adapted from Gries' deviation of proportions and degree of dispersion (DP) tests. Gries, “Dispersions and Adjusted Frequencies.” [↑](#footnote-ref-92)
93. Gesenius, Kautsch, and Cowley, *Hebrew Grammar*, §126w. [↑](#footnote-ref-93)
94. Julius Ley, “Über Den Gebrauch Des Artikels in Der Rhythmischen Poesie Der Hebräer,” *Neue Jahrbücher Für Philologie Und Paedagogik* 144, no. 2 (1891): 341–451. As cited by James Barr, “‘Determination’ and the Definite Article in Biblical Hebrew,” *Journal of Semitic Studies* 34, no. 2 (1989): 328–29. [↑](#footnote-ref-94)
95. Peter Bekins, “The Omission of the Definite Article in Biblical Poetry” (SBL Annual Meeting, San Antonio, 2016), 5n10, https://www.academia.edu/30145121/The\_Omission\_of\_the\_Definite\_Article\_in\_Biblical\_Poetry; Peter Bekins, “Non-Prototypical Uses of the Definite Article in Biblical Hebrew,” *Journal of Semitic Studies* 58, no. 2 (October 1, 2013): 233n14, https://doi.org/10.1093/jss/fgt001. [↑](#footnote-ref-95)
96. "If, for example, as Ley maintains, they treated the poetical literature by the rules applicable for prose, why did they not do this more consistently? How did a process of revision, even an unconscious one, lead to the numerous inconsistencies some of which are listed, for instance, by Lambert?" Barr, “‘Determination,’” 330. [↑](#footnote-ref-96)
97. "Is there a path from the deictic foundation of the article towards an explanation? For the moment I am content to leave it here." Barr, 331. [↑](#footnote-ref-97)
98. Humphrey Hardy and Samuel Boyd, “Hebrew Adverbialization, Aramaic Language Contact, and Mpny ʾšr in Exodus 19:18,” in *Semitic Languages in Contact*, ed. Aaron Butts (Brill, 2015), https://doi.org/10.1163/9789004300156; Gesenius, Kautsch, and Cowley, *Hebrew Grammar*, §117.1. [↑](#footnote-ref-98)
99. The Loca sample size N=13 in poetry means there is not enough data to make a definitive assessment for that genre. [↑](#footnote-ref-99)
100. Haspelmath, *From Space to Time*. [↑](#footnote-ref-100)
101. Bekins, “Non-Prototypical Uses,” 226–27. [↑](#footnote-ref-101)
102. Bekins, 226–30. [↑](#footnote-ref-102)
103. Bekins, 227. [↑](#footnote-ref-103)
104. Though only 25% with Time, it is the second-most frequent pattern and it also appears as a sub-construction within the most common pattern, definite attributive apposition, e.g. בַיוֹם הַהוּא. [↑](#footnote-ref-104)
105. "The major cyclic events of the human natural environment on earth have probably always served as the main means of locating and measuring other situations: in particular, the alternation of light and dark, changes in the shape of the moon, and changes in the path of the sun across the sky (accompanied by marked climatic differences)." Haspelmath, *From Space to Time*, 25. [↑](#footnote-ref-105)
106. While this analysis extends beyond the bounds of this article, the results of the preliminary experiment and the tagging itself can be viewed in the project repository linked in footnote 32. [↑](#footnote-ref-106)
107. As noted by Klein, *Time*, 150–58. [↑](#footnote-ref-107)
108. Robert L. Allen and Clifford A. Hill, “Contrast between ø and *the* in Spatial and Temporal Predication,” *Lingua* 48, no. 2–3 (June 1979): 123–46, https://doi.org/10.1016/0024-3841(79)90002-0. [↑](#footnote-ref-108)
109. Allen and Hill, 126–27. [↑](#footnote-ref-109)
110. Adapted from Allen and Hill, 135–38. [↑](#footnote-ref-110)
111. Allen and Hill, 135. [↑](#footnote-ref-111)
112. Allen and Hill, 140–41. [↑](#footnote-ref-112)
113. Allen and Hill, 138. See also Croft's remarks: "...there does seem to be a semantic parallel between the category of definiteness in nouns...and modal or modality related categories in verbs and complex sentence constructions such as world-creating predicates and conditional constructions." Croft, “A Conceptual Framework,” 265. [↑](#footnote-ref-113)
114. Exodus, in particular, appears to reflect more variability with בֹּקֶר and עֶרֶב when combined with עַד or מִן. See Exod 12:10, 12:10, 12:22, 16:19, 16:20, 16:23, 16:24, 18:13, 18:13, 18:14, 18:14, 27:21, 27:21, 29:34, 34:2. [↑](#footnote-ref-114)
115. The issue of the definite article in Hebrew time adverbials is complicated by a demonstrative use in time adverbials, e.g. היּוֹם 'this day'. Van der Merwe, Naudé, and Kroeze, *BHRG*, §24.4.4. [↑](#footnote-ref-115)
116. For "anchors" and "landmarks" see Janet Harkness, “Time Adverbials in English and Reference Time,” in *Essays on Tensing in English*, ed. Alfred Schopf, Linguistische Arbeiten 185, 228 (Tübingen: Max Niemeyer Verlag, 1987), 81; Charles Fillmore, “Mini-Grammars of Some Time-When Expressions in English,” in *Complex Sentences in Grammar and Discourse: Essays in Honor of Sandra A. Thompson*, ed. Joan L. Bybee, Sandra A. Thompson, and Michael Noonan (Amsterdam: Benjamins, 2002), 38. [↑](#footnote-ref-116)
117. E.g. Gesenius, Kautsch, and Cowley, *Hebrew Grammar*, §100c; Waltke and O’Connor, *IBHS*, 10.2.2c. [↑](#footnote-ref-117)
118. Haspelmath labels unmarked accusative adverbials which express a time span as "atelic extent." English is actually an exception in this case, preferring to use "for" to indicate such durational spans. Haspelmath, *From Space to Time*, 120–30. [↑](#footnote-ref-118)