11^{th} International Conference on Computational Social Science IC^2S^2
July 21-24, 2025, Norrköping, Sweden

Freezing and Thawing of Linguistic Binomials

Keywords: binomials, computational linguistics, language change, linguistic biases, temporal analysis

Extended Abstract

A *binomial* is any phrase consisting of two words separated by ‘and’; it has the interesting property that the order of the words must be determined solely by the speaker, not grammar. For instance, while mostly everyone says *bread and butter* instead of *butter and bread*, the binomial *Rachel and Tim* might have less agreement in the ‘natural’ ordering. Binomials with high agreement are called *frozen binomials* and have more than a century of study by linguists mostly in the form of snapshots or case studies [4, 1, 5, 7, 2]. However, work that explored the dynamic nature of binomials is relatively rare and most research has focused specifically on gendered terms without analyzing the overall dynamic patterns of binomials [9, 3]. One exception to this is [6] which similarly examined long-term shifts in binomials, using a small corpus of Google Books scans.

We build a corpus of binomials from 200 years of published American English work in the extensive HathiTrust Library so as to perform a large-scale analysis binomials over time. Our analysis challenges a number of long-held beliefs, particularly about binomials as a static property. Our work therefore seeks to shift the question from frozenness as a stable property that needs to be understood on static snapshots to frozenness as a highly dynamic property that varies over long periods of time.

Historical Frozen Binomials One interesting aspect of these prior works, is that many of these papers specifically identify binomials the authors believe to be frozen. Analyzing binomials identified as frozen from papers from 1920-1960, we have 125 binomials, with a few overlapping between publications. Of those, we have some data for 78 of them from our dataset, but only about 43 could plausibly be said to be frozen at the time of publication, even without taking into account the dynamic questions we consider here. That is, only about about half of binomials that previous work claimed to be frozen would be classified as frozen by our standards at the time they were publicized. Since these works relied on author discretion to identify frozen binomials, this could suggest that the common academic understanding of ”frozenness” does not fully align with the actual usage of binomials, and indicates that the conceptual foundations of binomial analysis are not fully settled.

Freezing and Thawing Our analysis challenges some long-held beliefs about binomials. First, it is common belief that binomials, once frozen, rarely thaw. We find, in fact, the opposite. Thawing is far more common than freezing (Fig 1). This is possible, in part, because of another historical claim that we do find support for. Frozen binomials are used more frequently than non-frozen binomials. Thus, many binomials ‘appear’ in our dataset frozen (when they first acquire a rate of usage high enough to be analyzed), and then thaw later. However freezing is a relatively quick process - most binomials freeze in less than 50 years. Thawing can take many decades, or even more than a century. However, we also see far more instances of thawing than freezing in our dataset. This is due to the relative frequency of frozen binomials.

Trimodal Ordinality A key definition in our work is the *ordinality* of a binomial: the fraction of time it appears in alphabetical order. Thus, ordinalities close to 0 or 1 means that binomial is almost always written in the same order, indicating that it is frozen or nearly frozen.

(Alphabetization is not crucial for this definition; any canonical ordering on words would be sufficient for us in identifying binomials whose order is nearly fixed.)

The histogram of the ordinality of all binomials reveals an interesting structure (Fig 2). About 30% of common binomials are considered frozen, and exist on either end of the ordinality spectrum. The rest, are normally distributed around 0.5. This trimodal structure is unusual; while it might suggest a fundamental divide between frozen and non-frozen binomials, given the apparent dynamism of binomials discussed earlier, there is no evidence of this divide. Instead, we ask if there is a simple model that could explain this structure. Here, we introduce one such model that we refer to as $k - peeks$ (see Algorithm 1). Roughly speaking, the $k - peeks$ framework is a stylized model in which users look at the last k instances of a binomial they have seen, and randomly choose one of them to copy the order. This model, with tuning, can reproduce the trimodal structure.

Topical Changes Most of the changes we see related to previously identified binomials seem related to slow, long-term changes. Two common examples of long-term cultural shifts present in binomial orderings are gender-related binomials and national binomials. Gendered binomials consist of two words that are explicitly gendered such as ‘brother and sister’ or ‘steward and stewardess’. A common conclusion across most previous works is the observation that binomials tend to be ordered male first. In fact, this is one of the oldest ’rules’ presented across the field. When we look at gendered binomials temporally, we see that this rule does seem to be true, but it is also changing (Fig 3). If we look at familial terms over 200 years, it appears that many of these terms have thawed from exclusively male-first. Most of these trends are long-term, with the rate of change almost constant for the duration of our data. While the magnitude of the change varies from binomial to binomial, the direction is always the same.

We test national binomials by gathering lists of countries and finding binomials that contain only country names. We find two major results. First, there are strong but shifting hierarchies of countries. That is, if A comes before B and B before C, then A comes before C in most cases. Despite these statistically significant hierarchies, national binomials are relatively mobile. In the short term, sudden changes sometimes appear around major international incidents such as wars. While these changes are often relatively small, they are significant compared to the amount of shift normally found in binomials. The language is also ’americanizing’ (Fig 4). America is becoming first more than it’s becoming second in binomials. Out of national binomials containing the word ’america’ or ’american’, respectively, 4 became less americanized, 21 more americanized, and the rest (35) did not significantly change.

Future Work In this work we explore binomial evolution through a new dataset of historical binomials involving two centuries of text. This large dataset and long time frame allows us to explore novel aspects of binomials: the freezing and thawing patterns and the trimodality of ordinality. Establishing the dynamism of binomials opens up new avenues to an old topic. The oldest question of binomials is why some are frozen and some are not - and why they are frozen in the direction they are. This question has gone largely unanswered; however, leveraging the changes in binomials alongside changes in pronunciation, importance, meaning, and other external factors allows for an entirely new approach to our understanding of binomials. By understanding how and when binomials change, we can begin to understand how they came to be the way they are. Additionally, while the present project focuses on American English, binomials have been studied in many languages and it would be useful to reproduce these results across languages. Further, it has been previously established that language models have absorbed binomial biases from their text [8], but additional work is needed to understand their reproduction of these biases - and how quickly models can respond to changing norms.

References

[1] Richard D. Abraham. Fixed order of coordinates: A study in comparative lexicography. *The Modern Language Journal*, 34(4):276–287, 1950.

[2] William E. Cooper and John Robert Ross. Word order. In Robert E. Grossman, Jame L. San, and Timothy J. Vance, editors, *Papers from the Parasession on Functionalism*, chapter 6, pages 63–111. Chicago Linguistic Society, Goodspeed Hall, 1050 East 59th Street, Chicago, Illinois 60637, 1975.

[3] Peter Hegarty, Nila Watson, Laura Fletcher, and Grant McQueen. When gentlemen are first and ladies are last: Effects of gender stereotypes on the order of romantic partners' names. *British Journal of Social Psychology*, 50(1):21–35, 2011.

[4] Otto Jespersen. *Growth and Structure of the English Language*. B.G. Teubner, 1905.

[5] Yakov Malkiel. Studies in irreversible binomials. *Lingua*, 8:113–160, 1959.

[6] Sandra Mollin. Pathways of change in the diachronic development of binomial reversibility in late modern american english. *Journal of English Linguistics*, 41(2):168–203, 2013.

[7] Steven Pinker and David Birdsong. Speakers' sensitivity to rules of frozen word order. *Journal of Verbal Learning and Verbal Behavior*, 18(4):497–508, 1979.

[8] Katherine Van Koevering, Austin R Benson, and Jon Kleinberg. Frozen binomials on the web: Word ordering and language conventions in online text. In *Proceedings of The Web Conference 2020*, pages 606–616, 2020.

[9] Saundra K Wright, Jennifer Hay, and Tessa Bent. Ladies first? phonology, frequency, and the naming conspiracy, 2005.

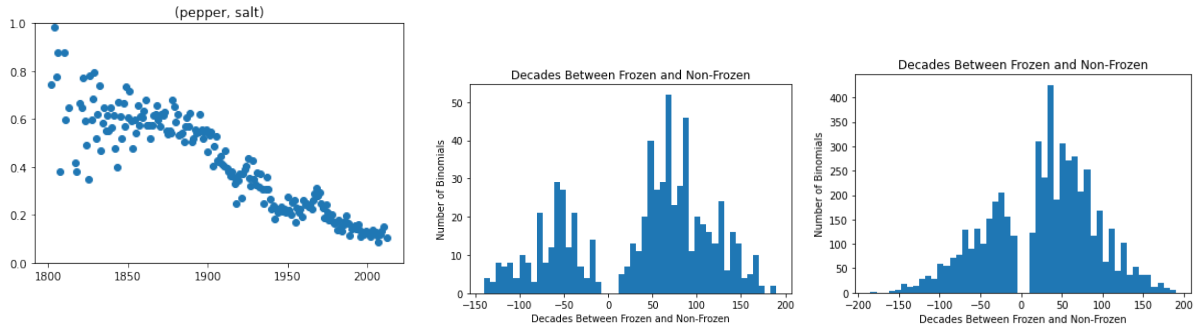

Figure 1: **Dynamism of Binomial Frozenness**(A) An example of a dynamic binomial. The binomial (pepper, salt) first appears in our dataset in the year 1800. Due to sporadic data, the true ordinality is unclear, but there is a preference for the order *pepper and salt*. By 2000, that preference has flipped and the binomial is now frozen *salt and pepper*. (B) For all binomials that have been both frozen and non-frozen, a histogram of the number of years between their first year frozen and their first year non-frozen. On the left, we define non-frozen as ordinality greater than 0.45 or less than 0.55, on the right non-frozen is defined as asymmetry less than 0.75 or greater than 0.25. However, we additionally only consider a binomial frozen or non-frozen if it remains so for at least 15 years. For these plots, we relax our frequency requirements to 10 instances per 5 years, as opposed to 30. Adjusting these parameters has little effect on the conclusions.

K-Peeks Model

For each binomial, we assign orders to some n timesteps of instances as follows.

```
Data: $k > 0$ window size, $n > k$ time steps, and $0 \leq p \leq 1$ probability
$t \leftarrow 0$;
while $t < n$ do
    $i \leftarrow U(0,1)$;
    if $i \leq p$ then
        $j \leftarrow U(0,k)$
        $t+1$ order $\leftarrow$ order from $t-j$ timestep;
    else
        $j \leftarrow U(0,1)$
        if $j < 0.5$ then
            $t+1$ order $\leftarrow$ alphabetical order;
        else
            $t+1$ order $\leftarrow$ anti-alphabetical order;
        end
    end
end
```

Algorithm 1: K-Peeks Model

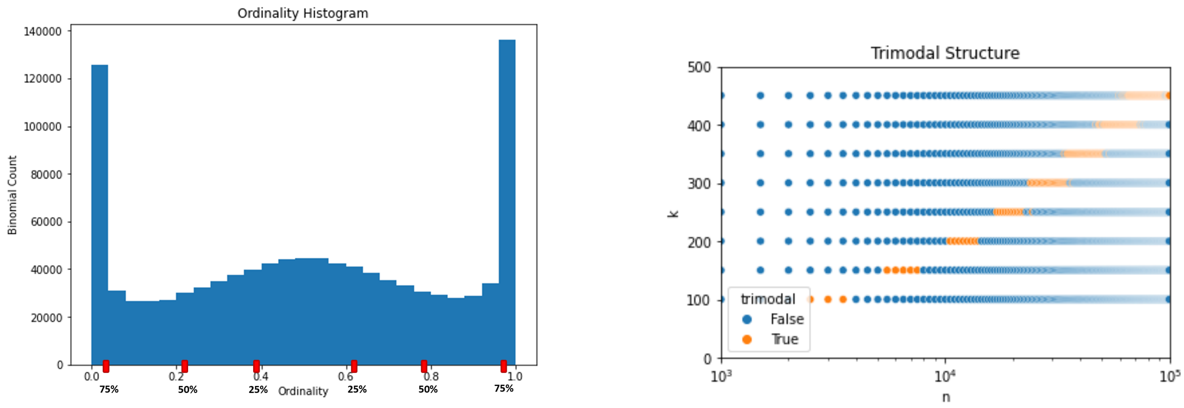

Figure 2: **Trimodal Structure of Ordinality**(A) Binomial ordinality follows a trimodal structure. There are a large number of frozen binomials (with ordinality at or close to 1 or 0). Non-frozen binomials tend to follow a roughly normal distribution centered at 0.5. This has also been demonstrated in prior work. The red lines demonstrate the data cut into roughly equal pieces. (B) Given $p = 0.998$, the states for n and k where trimodal structure is present in the histogram of ordinalities. Note that the x-scale is logarithmic and the relationship is close to linear.

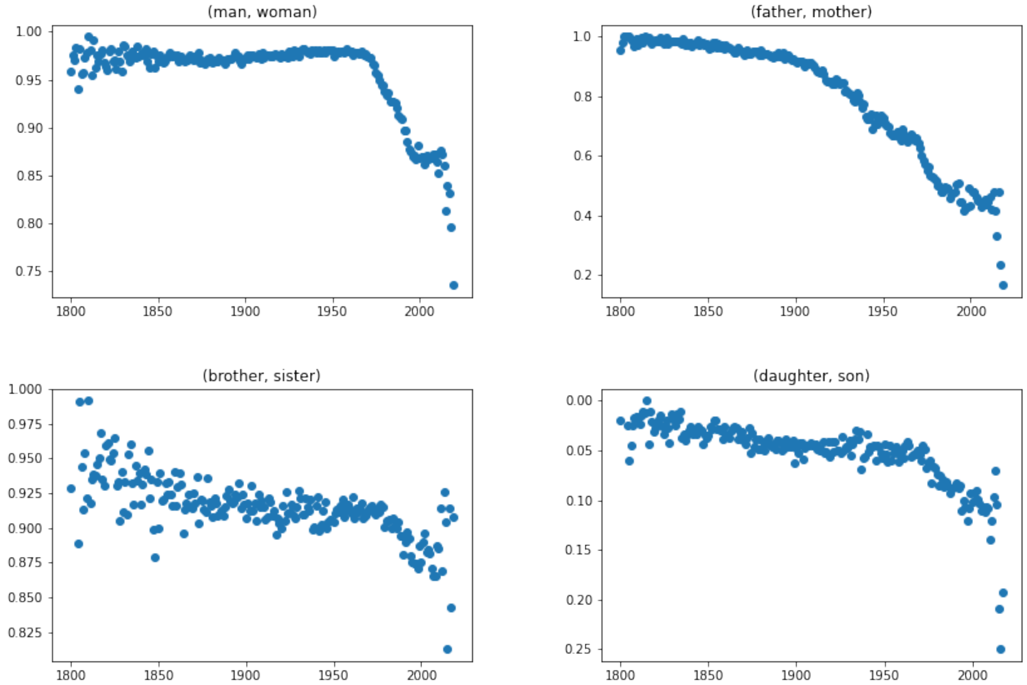

Figure 3: **Gender Dynamism** We examine four of the most popular family-term binomials. For each, we include both the words and their plurals. Note that the terms universally begin frozen male-first, but then tend towards female first over time. The rapidity of the trend varies, but most of the change occurs post 1950. Note that we flipped the axis of (daughters, sons) to maintain the 'downward trend' similarity of the other plots.

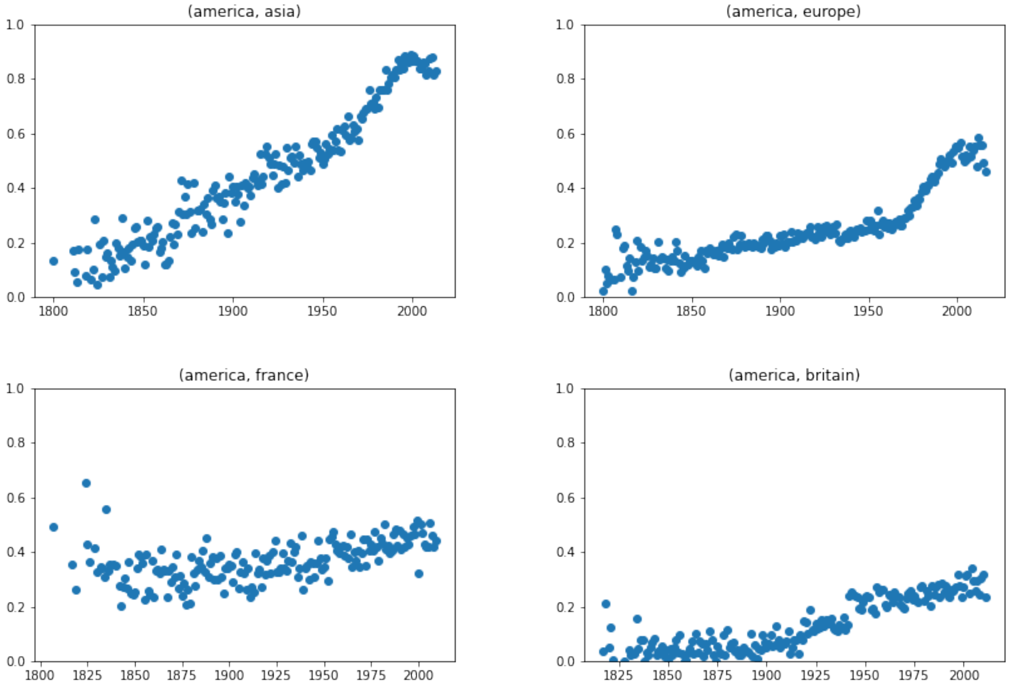

Figure 4: **Nationality Ordinality** These plots show the oridinality shift for several of the most common binomials with 'America' as one word. Note that for all of them, the long-term trend is towards 'America' first. This could represent the overall 'Americanization' of the American English language, as the country (and continents) gain more power on the global stage and more presence in the minds of the authors.