THE EXPRESSION OF THE MANNER OF MOTION IN PORTUGUESE AND OTHER LANGUAGES: A BAYESIAN STATISTICAL ANALYSIS OF ORAL NARRATIVES

OLIVEIRA, Aparecida de Araújo[1*]
FERNANDES, Rafael Macário[2]
CARVALHO, Luiz Max[3]

[1] Federal University of Viçosa – ORCID: https://orcid.org/0000-0002-8888-6496
[2] University of São Paulo – ORCID: https://orcid.org/0000-0002-6742-3206
[3] School of Applied Mathematics, Getulio Vargas Foundation – ORCID: https://orcid.org/0000-0001-5736-5578

Abstract: *We compared verbs expressing the manner of motion in oral narratives by speakers of Romance and Germanic languages, respectively classified as verb-framed and satellite-framed according to L. Talmy's Motion Typology. We used a Bayesian statistical analysis on aggregated and disaggregated data from 12 speakers per language. For Brazilian Portuguese, we used data from 14 speakers. Considering inter-typological variation, the manner per motion-verb ratios demonstrated a more varied use of manner verb types by the speakers of the satellite-framed languages, as expected from theory: German (0.33), English (0.46), Spanish (0.18), Italian (0.17), French (0.08), and Brazilian Portuguese (0.23). These results also show a substantial similarity between Brazilian Portuguese, Spanish, and Italian, with slightly fewer types of manner verbs in the French narratives. In general, speakers of Germanic languages preferred to use manner over other kinds of verbs, with mean per-speaker rates of 5.10 and 5.27 for German and English, respectively. In contrast, the mean rates demonstrated a preference for other verb types (such as path verbs) among speakers of Romance languages: Italian (1.37), French (1.22), Brazilian Portuguese (2.21), and particularly Spanish (0.98): Analysis of disaggregated data for Brazilian Portuguese and English showed substantial variability among speakers in their use of motion verbs in general and manner verbs in particular. It reduced the rate of manner-verb usage in English to 4.35 and in Brazilian Portuguese to 1.86.*

Keywords: Verb semantics; Talmy's Typology; Romance languages; Germanic languages; Bayesian statistics.

[*] Corresponding author: idaaraujo007@gmail.com

1 Introduction

In his hypothesis known as Motion Typology, cognitive semanticist L. Talmy (1, 2) proposes grouping languages according to two lexicalization patterns that correspond to characteristic ways of mapping particular semantic elements into linguistic forms. Considering the grammatical resources each language provides its speakers with to express the path of motion, he places Romance and Germanic languages on opposing sides. However, these idioms also differ regarding the lexicon of verbs that denote motion. Some languages have more verbs that encode the path, and others have more verbs that encode the manner of motion.

The manner in which a moving figure goes about is a secondary element in the general constitution of motion events and may be left unspoken according to a speaker's necessity and language type (2, 3). Still, it has aroused significant interest regarding how speakers of various languages typically express it if they choose to do so: for instance, Phelps and Duman (4) comparing English and German; Özçalışkan and Slobin (5) about English, Spanish, and Turkish; Hijazo-Gascón and Ibarretxe-Antuñano (6) about Spanish, Italian, French, and German, besides a myriad of works by Slobin and his followers, on both native and second or foreign languages. As native speakers of Brazilian Portuguese, we have become interested in investigating the expression of motion events in our language, with a focus on the manner component in the verb root. We thus compare the expression of the manner in verbs used to describe motion events in Brazilian Portuguese, English, Spanish, Italian, French, and German in the face of L. Talmy's hypothesis.

Talmy's Motion Typology has been widely explored in connection with D. Slobin's (7) Thinking for Speaking Hypothesis, for which language does not affect general thinking but interferes with how we think as we *speak* about our experiences and pay more attention "to the dimensions of experience (...) enshrined in grammatical categories." (8, p71). Despite its longevity after its recognition as a groundbreaking theory, Talmy's two-category typology has not been undisputed, as shown below. With that in mind, our primary goal was to compare Brazilian Portuguese to the languages in the Germanic and the Romance groups, which means we sought inter-typological and intra-typological variation (9) in the expression of the manner of motion. Our second aim was to demonstrate how a Bayesian statistical method allowed us to use our disaggregated data and the aggregated data from (6) to include Portuguese among *"las lenguas románicas"* Hijazo-Gascón and Ibarretxe-Antuñano investigated.

2 A typology of lexicalization of motion events

Lexicalization patterns are typical linguistic ways to encode meaning components into words and syntactic constructions. They are recurrent in everyday speech and pervasive in particular domains (Talmy, 1, 2). Lexicalization patterns explain static and dynamic situations, but we restrict our analysis to events with a shift in an entity's position in space. The examples below describe a scene in Mayer's frog story (10), retrieved from English and Brazilian-Portuguese narrative corpora.

(1) The frog **tiptoes**_{verb.MANNER} **out**_{satellite.PATH} **of** the jar_{PP.GROUND}. (Eng_Slobin_20k) > S-language

(2) *O sapo **sorrateiramente***_{adverb.MANNER} ***sai***_{verb.PATH} ***do pote***_{PP.GROUND}. (Port_2018_10) > V-language
[Lit.: The frog surreptitiously exits the jar.]

Talmy (2, 11, and elsewhere) uses the term 'conflation' to refer to the fusion of semantic components in a verb root. From the examples above, the English verb *tiptoe* conflates a general

idea of motion and how it occurs: moving quietly or cautiously on or as if on tiptoe. [1] Conversely, the Portuguese verb *sair* (exit) conflates motion and an outward direction or path. Finally, a verb like *snow* conflates the idea of movement and the figure (snowflakes).

Verb semantics defined Talmy's (12) language typology. Here, though, we use its 1991 version, which categorizes languages according to which forms they use to encode the path element. Example 1 is a typical case of the satellite-framed lexicalization pattern found in many languages worldwide, including those of the Germanic family. We call these S-languages as they express the general direction or path of motion outside the main verb in a class called a "satellite." Satellites are particles such as the English *out* and the German *aus* or affixes in other languages. Example 2 illustrates the verb-framed lexicalization pattern observed in Romance and many other language families, which we refer to as V-languages. In Portuguese, for instance, the verb *sair* (to exit) encodes the path (or direction) of the moving character, called the figure, against the scenery, named *ground.*

As the examples above demonstrate, a motion event comprises additional semantic components beyond the path. Talmy also identifies the manner, the cause, and other elements as distinct categories of co-events. [2] S-languages usually have a manner verb as the predicate head, for instance, *tiptoe* in Example 1. At the same time, V-languages tend to express this component outside the predicate, in adverbial phrases such as the Portuguese *sorrateiramente* in 2.

3 The expression of the manner

The manner occurs elsewhere in the clause other than in predicate heads and adverbials, such as gerunds of manner verbs, as seen below:

> (3) … *foram **correndo**_{＜gerund.MANNER＞} atrás dele… as abelhas.* (Port_2018_2) > V-language
> [Lit.: … went running after him … the bees.]

S-language speakers are more prone to express the manner when describing motion scenes. On the other hand, because the manner is a supporting event, its materialization in discourse is optional in V-languages. One must not take optionality as random, however. One can attribute it to at least two factors.

First, as mentioned earlier in this paper, the grammar of a language is assumed to limit what speakers select to say. According to Talmy (2), this is a matter of salience, which he formalizes in four principles (see p128–30). In short, if a semantic component is part of the verb root or a closed-class form, such as a satellite, it is backgrounded, i.e., speakers will express it effortlessly because it is inherent to the verb complex. This situation is typical in S-languages regarding the expression of the manner because it appears mainly in the verb root.

In contrast, if a semantic component appears outside the verb complex, it is foregrounded, meaning speakers are more aware of its presence in their speech. As a result, its expression requires a more significant cognitive effort, often leading to its omission from discourse. Such is the case of V-languages concerning manner expression because this lexicalization pattern predicts that manner will emerge in gerunds and adverbials, for instance, as mentioned above.

A second point is that, besides the distinctive distribution of grammatical elements, researchers have long accepted that S-languages have a more extensive lexicon of verbs that conflate manner and motion, such as *crawl*, *swoop*, and *tumble* (see 13), which probably allows speakers to attend more to the manner of motion. Accordingly, these verbs constitute two

---

[1] The Merriam-Webster.com Dictionary ®

[2] For a full list of the co-events, see Talmy (1, p484) and (2, p49)

recognizable "tiers" in terms of volume and diversity, differing according to the language type, as seen by Slobin (14, p458). Verbs of the first tier, such as *fly*, *run*, *walk*, and *climb*, express fundamental distinctions and find correspondence in most languages. The second tier includes verbs such as *dash* and *scramble*, which are deemed more "expressive" or "exceptional" and mark manner distinctions in S-languages that V-languages do not easily convey in verbs.

Moura and Badaracco (15) argue that English speakers also benefit from a morphological resource called zero-derivation to increase their means of expressing the manner in the main verb. Such a highly abstract word-formation process allows lexemes to assume a different word class depending on the context in which they occur (16). For instance, the zero-derived verb *wrist* arises from a noun in the appropriate context, as in Example 4, where the manner might emerge as an instrument in the movement.

(4) He **wristed** the ball over the net. (15, p169)

Taking a different stand, and based on data from former Twitter, Santos Filho (17) argues that Brazilian Portuguese has numerous verbs that encode fine manner traits and can be the head in a clause. According to him, around 120 of these verbs conflate motion and manner, nearly 100 express motion, path, and manner, and a few encode motion and figure. His findings are relevant because they emerge from the analysis of colloquial-style data and show several "expressive" verbs. However, using these verbs is subject to the boundary-crossing constraint identified by Aske (18), who observed that V-languages do not allow manner verbs as the head of the predicate to describe movement *out of*, *into*, or *across* a place. This constraint does not exist in S-languages.

So, while S-Languages typically encode the manner in the main verb, Romance languages, which have fewer manner verbs, express this semantic element in an independent component, perhaps a gerund following a path verb, an adverbial phrase, or a prepositional phrase, and, thus, may even omit it altogether, as mentioned above (12).

These observations certainly strengthen Talmy's typology. However, when researchers applied it to more idioms, they encountered facts contradicting his binary categorization. As a result, besides looking for intergroup differences that partially support Talmy's typology or beg for more categories, they also seek variations across languages in individual groups.

4 Intergroup and intragroup similarity and variation

Evidence-based criticism has emerged regarding the categorical attribution of languages such as Chinese Mandarin, which Talmy considers akin to English as a motion-plus-manner language (1, 12). Zlatev and Yangklang (19) propose that Thai, Mandarin, and other serial-verb languages constitute a third type, distinct from V-languages and S-languages since they have typical verb constructions expressing manner, path, and deixis in individual verbs in a single clause, as seen in Slobin and Hoiting (20). From these findings, Slobin (3) proposes a third type, equipollently framed, an idea Talmy (21) questions with a series of counterexamples. Still, Talmy's classic illustration of "the bottle floating out of the cave" differentiates English from Spanish, a major distinction between Germanic and Romance languages that is now unequivocally acknowledged.

Researchers have also sought variation among languages within the same group. This latter is known as *intra-typological variation* (9, p533) and corresponds to diverse "levels of salience and granularity of motion semantic components" found in languages of the same type, such as Spanish and Basque, that are of different families, but also between languages of the same family, such as the Romance languages Spanish, Italian, and French. As a result, some authors have proposed a salience *cline* of semantic components instead of a set of distinct groups, as in Talmy's typology (see, for instance, 3).

Our study uses Hijazo-Gascón and Ibarretxe-Antuñano's (6) findings about inter-typological and intra-typological variation. First, as expected for an S-language, the German data they analyzed yielded many verb types and other structures that conveyed manner. They found little or no variation in the expression of manner among Romance languages, much less outstanding than the variation in the expression of the path. Similarly to what one expects for V-languages, Spanish, French, and Italian speakers paid little attention to the manner, having hardly ever expressed it in the main verb or other forms as in German. The manner-verb types they used were few, the majority conveying basic manner senses, such as those meaning *run,* and even fewer verbs of the second tier.

They also referred to Italian syntagmatic (or analytic) verbs as a unique distinction from French and Spanish. For instance, *scivolare fuori* (6, p473), meaning *slip out*, would translate into Portuguese as *escapar*, a path verb. Still, they were reluctant to regard Italian as an S-language or a language with medium salience for manner. On the other hand, they considered French a V-language in general because it does not have idiosyncratic elements like Italian and confirmed Spanish as a typical V-language in all senses. While Hijazo-Gascón and Ibarretxe-Antuñano did not obtain significant evidence for intra-typological variation in the lexicalization of the manner among Romance languages, their findings supported their expectations concerning a significant intergroup variation when contrasting these with German.

5 Our study

Focusing on the expression of the manner of motion, we set our objectives as follows: our primary goal was to add Brazilian Portuguese to the Romance languages investigated in (6) to seek intra-typological variation. We also found that comparing German to another Germanic language was suitable, so we brought in available raw English data. Our third purpose was to seek inter-typological variation by comparing our results for Brazilian Portuguese with theirs about German and ours for English. Finally, we searched for contrasts across the two groups of languages sampled here.

Such comparisons were possible because we obtained all our data using the same method. That enabled us to quantitatively analyze the same scenes as Hijazo-Gascón and Ibarretxe-Antuñano for all six languages.

5.1 Materials and Methods

Our study used both direct and indirect data obtained through the well-known frog-story method created by Bamberg (22) [3] and employed in numerous works about this and other themes up to this date, such as Berman and Slobin (23), Slobin (3), and Guo et al. (24). See also Ibarretxe-Antuñano (25). We analyzed Brazilian-Portuguese and English raw data in comparison to indirect data from Hijazo-Gascón and Ibarretxe-Antuñano's frog-story study, which included Spanish, French, Italian, and German.

Data collection involved eliciting oral narratives based on a wordless picture book titled "*Frog, Where Are You?"* (10). In this 24-scene story, the protagonists — a boy and his dog — engage in various motion situations as they search for a fugitive frog. We chose the clause as our unit of analysis and identified the path and manner components, following Talmy's (1) categories.

We used disaggregated data for Brazilian Portuguese and English. Our Brazilian-Portuguese data consisted of a sample from the project *Distinct Grammars and Their Role in the*

---

[3] For a complete description of Bamberg's method, access the *Childes Talk Bank*, at https://childes.talkbank.org/access/Frogs/0docs/0frogs.html.

Acquisition of L2 Portuguese (26). Participants were college students at the Federal University of Viçosa, in southern Brazil, with an average age of 22.56 years, raised as monolingual speakers of this language. The 12 English narratives were elicited from 20-year-old students at Berkeley, California, by Marchman (27) as part of Berman and Slobin's (23) crosslinguistic project about language development, and are available at the Childes Talkbank project.[4]

Lastly, we used aggregated data from (6) about Spanish, Italian, French, and German. Although we were unable to access their per-participant data, we could still compare our results with theirs by employing the Bayesian statistical analysis described below. Their original data also consisted of frog stories ($N_1 = 12$) elicited from native speakers of each language, aged, on average, 21.58 years (Spanish), 23.91 years (German), 21.67 years (French), and 21.58 years (Italian) (p476).

5.1.1 A Bayesian statistical analysis

We used the R language for our statistical procedures, utilizing the relevant packages detailed in the Appendix. All analyses were conducted on a Dell G5 with an Intel Core i7 processor and 16 GB of RAM, running GNU/Linux Ubuntu 22.04. Data and code can be found in [32].

Choosing the prior distribution for a model's unobserved quantities ("parameters") is crucial in a Bayesian analysis. Here, we decided on a Normal (0,5) prior for the intercepts and Normal (0,1) (standard normal) priors for the random effects. We parametrize the normal distribution by mean and standard deviation. These are common prior choices in Bayesian literature, called weakly informative priors. They lead to well-behaved inferences without appreciably impacting the results. See Gelman et al. (28) for more details.

We addressed these questions about the "manner verbs":

I – Do languages differ in the ratio of manner verbs? Here, we examined the relation between the total number of motion-verb and manner-verb types.

II – Do languages differ regarding the mean rate of manner verb occurrence?

III – Do languages differ in the mean rate of occurrence of manner verbs, taking inter-speaker variation into account? Here, we should be able to tackle the different production levels of manner verbs across speakers.

We expected to find more distinctions between the Brazilian Portuguese and the English-German pair. Supposedly, but not necessarily, there would be a substantial degree of similarity across the four Romance languages and also some variation between these.

5.2 Results and Discussion

Our analysis of the data provided the following verbs, which conflate solely the motion and the manner, as in Hijazo-Gascón (29, p361) and Hijazo-Gascón and Ibarretxe-Antuñano (6, p477). Considering the numbers after each verb as the number of tokens, for Brazilian Portuguese we observed: *correr* (run) (23), *pular* (jump) (4), *carregar* (carry) (2), *andar* (walk) (1), and *arrastar* (drag) (1). The number of path verbs surpassed that of manner verbs: *chegar* (arrive) (1), *deixar* (leave) (1), *descer* (descend) (2), *esconder-se* (hide) (3), *sair* (exit) (25), *partir* (leave) (1), *seguir* (follow) (2), *subir* (ascend) (16), and *voltar* (go back) (1). Likewise, the number of tokens of manner+path verbs in this language also outmatched that of the manner-only verbs: *escapar* (escape) (4), *cair* (fall) (51), *fugir* (flee) (11), and *perseguir* (chase) (6). Neuter verbs were few in

---

[4] Available from https://childes.talkbank.org/access/Frogs/English-Slobin.html.

type but also many in occurrences: *ir* (go) (33) and *levar* (take) (5). Finally, a single verb type, *parar* (stop) (5), conveyed only the movement's final part.

As we expected for Brazilian Portuguese, the low number of manner verbs is a characteristic of V-languages compared to path verbs and other non-manner ones. Still, a few remarks are due regarding the imbalance in the number of tokens among the verb types. Firstly, the high frequency of the verbs *correr*, *subir*, and *cair* has a parallel in the English verbs *run*, *climb*, and *fall*, which one can easily attribute to the number of scenes of running, climbing, and falling in the story, and to their typicality in everyday use in both idioms. The verb *sair also* has a high rate of occurrence due to the three scenes of going out and leaving. This verb may also have an exclusively aspectual function, such as in Example 7, but we did not count such occurrences of *sair*. Finally, as typical with V-languages, the majority of the occurrences of the manner verbs in Brazilian Portuguese were in the gerund or infinitive form, as in 7-8.

(5) *o animal sai do buraco.*
[Lit.: the animal comes out of the hole.]
(6) *saíram para procurar o sapo pela floresta.*
[Lit.: '(they) left to search for the frog in the forest.']
(7) *O antílope saiu correndo junto com o cachorro.*
[Lit.: 'The antelope set off running along with the dog.']
(8) *O cachorro começou a correr das abelhas*
[Lit.: 'The dog started to run from the bees.']

For English, on the other hand, we found far more types of only-manner verbs: *carry* (3), *crawl* (4), *creep* (1), *flap* (1), *fly* (4), *give.a.ride* (1), *hop* (3), *jump* (4), *limp* (1), *pop* (4), *run* (27), *slip* (1), *sneak* (2), *splat* (1), *step* (1), *swarm* (1), *get thrown* (2), *tiptoe* (1), *walk* (1), and *wander* (1). [5] According to our analysis, there were also nine types of path verbs in the English samples, which were *approach* (2), *come* (13), *depart* (1), *follow* (6), *hide* (1), *head* (2), *lead* (1), and *leave* (2). However, these correspond to around half of the uses of path verbs in Brazilian Portuguese and about half of the uses of the manner verbs in English. Alongside these verbs, the English participants also used *escape* (9), *chase* (15), *climb* (17), *fall* (34), *plummet* (1), *tumble* (2), and the neuter verb *go* (30). Again, we found a verb borrowed from Romance (*escape*) and a lesser amount of usage of *fall*, corresponding to approximately 65% of its counterpart, *cair*, in Brazilian Portuguese.

We expected such contrasts from the existing literature on Motion Typology. Moreover, we should also qualify the English path verbs that do not express very salient directions, such as those in Romance verbs that mean *ascend*, *descend*, *exit*, *enter*, and *return*. English accommodates these directions in satellites such as *up*, *down*, *out*, *in*, *back*, etc. Two of these path verbs — *approach* and *depart* — have a Latin origin through French. Finally, the most frequent path verb found in the English narratives — *come* — is, in fact, deictic.

So far, our findings about motion verbs in both languages have been consistent with theory. We now intend to examine the data quantitatively.

Table 1 summarizes the aggregated data for all languages concerning manner-only verbs, partially from (6) and partially from this study. As expected, a preliminary observation demonstrates a distance between English and German on one side and the four Romance

---

[5] For the verbs in the other languages, the reader can refer to (6).

languages on the other regarding the number of manner-verb types. Some variation is also visible within each group. We will address both issues below.

Table 1: Aggregated data on manner-verb use in six languages.

Language	Total types	Manner types	No. speakers	Occurrences (manner)
Spanish	39	7	12	11
Italian	60	10	12	16
French	53	3	12	14
Brazilian Portuguese	21	5	14	31
German	67	23	12	62
English	41	20	12	64

We also used disaggregated (per speaker) data: $N_1 = 14$ speakers for Brazilian Portuguese and $N_2 = 12$ speakers for English, as shown in Figure 1 below. Substantial variability is noticeable (cf. the y-axis) between speakers in their use of motion verbs in general and manner verbs in particular, as the dots about each small rectangle show. A proper statistical treatment will consider this variability when we assess between-language differences.
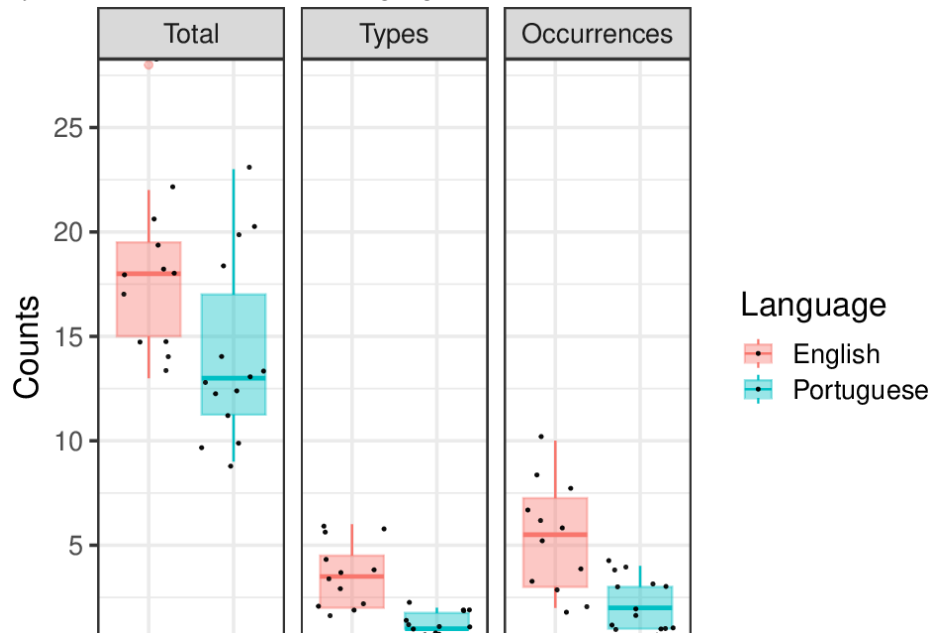
Figure 1: Boxplots of disaggregated counts for total types of motion verbs, types of manner verbs, and tokens of manner verbs.

We first analyzed the counts aggregated by language, ignoring within-language variation. The first question we tackled was whether languages differed in the ratio p_i of verbs that conflate motion and manner. In our notation, n_i represents the total number of motion verbs observed in a language i, and x_i represents the number of manner verbs among those. Our statistical model for this problem is the binomial distribution, where we parametrize the normal distribution in terms of mean and standard deviation: $x_i \sim$ Binomial (n_i, p_i); $p_i = 1/(1+\exp(-[\theta + \delta_i]))$; $\theta \sim$ Normal $(0,5)$; $\delta_i \sim$ Normal $(0,1)$.

Using a *random-effect* structure, we modeled the ratio, p_i, with an overall effect, θ, which captures the overall ratio of manner verbs across all languages, and a per-language effect, δ_i, which reflects the modification each language makes to the "grand mean" represented by θ. The

transformation $1/(1 + exp(-x))$ maps θ and δ_i to the space of probabilities $(0,1)$ and is called the *inverse logit* function (30).

　　We fitted our model to the aggregated data using Stan via the cmdstanr package (31), with four chains of 1,000 iterations (500 warm-up iterations, 500 sampling iterations) each. Details on the computing environment are available in the Appendix. Most runs took around 23 seconds in total (an average of 17 seconds per chain). We reported the posterior mean and 95% Bayesian credible intervals (BCI) for the estimates of the ratio p_i for each language. The results in Table 2 show Spanish and Italian with very similar manner-verb ratios per motion verbs and French with remarkably smaller ratios. Moreover, as expected by theory, the S-languages German and English show ratios much above the V-languages: 1/3 of all German motion verbs are manner verbs, and nearly half stands for English verbs.

Table 2: Posterior mean and 95% BCIs for manner-verb ratio.

Language	Mean	Lower	Upper
Spanish	0.18	0.09	0.31
Italian	0.17	0.09	0.27
French	0.08	0.03	0.15
Brazilian Portuguese	0.23	0.10	0.42
German	0.33	0.23	0.45
English	0.46	0.32	0.61

　　In Figure 2 and elsewhere in this paper, BP stands for Brazilian Portuguese. Here, the graph allows us to visualize patterns that might be hard to spot in the table. For instance, it becomes immediately apparent that although German and English have very similar ratios of manner verbs, the uncertainty around these values is higher for German.
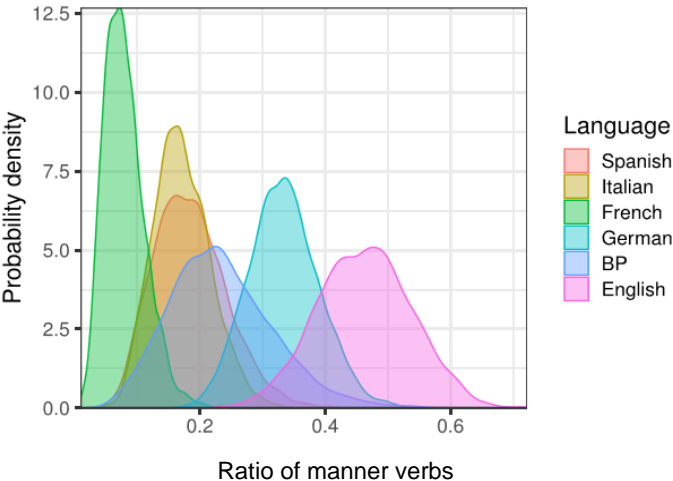
Figure 2: Posterior density of the ratio of manner verbs for each language.

　　To address question II, whether languages differed in the *rate* of occurrences of manner verbs, we used a Poisson distribution for the different numbers of occurrences. As there were 14 Portuguese speakers and 12 for the other languages, we accounted for this variation by using the n_i as an *offset* and estimated a *per-speaker* rate. In the equation below, y_i represents the total

number of times a manner verb was used across all speakers. We were then interested in the rate λ_i of occurrences of manner verbs in a language i, i.e., the average of manner verbs used per speaker.

This setup leads to a model that reads: $y_i \sim$ Poisson $(n_i \lambda_i)$; $\lambda_i = \exp(\alpha + n_i)$; $\alpha \sim$ Normal $(0,5)$; $\eta_i \sim$ Normal $(0,1)$. It is similar in spirit to the previous one. We have an overall (log) rate, α, and a language-specific effect η_i. Figure 3 and Table 3 show the results of fitting this model. As expected from theory, the S-languages German and English showed the highest manner-verb rates per speaker, while the Romance languages had lower rates, albeit with substantial variation. Brazilian Portuguese and Italian had the highest rates within their group, and Spanish had the lowest. The French showed a rate comparable to that of the Italians, with greater uncertainty. Again, the results show some non-significant variation across the Romance, as indicated by the overlapping green, red, and beige curves in Figure 3.

Table 3: Posterior mean and 95% BCIs for the rate of occurrence of manner verbs estimated from aggregated data.

Language	Mean	Lower	Upper
Spanish	0.98	0.52	1.58
Italian	1.37	0.82	2.09
French	1.22	0.68	1.91
Brazilian Portuguese	2.21	1.51	3.04
German	5.10	3.89	6.38
English	5.27	4.08	6.62
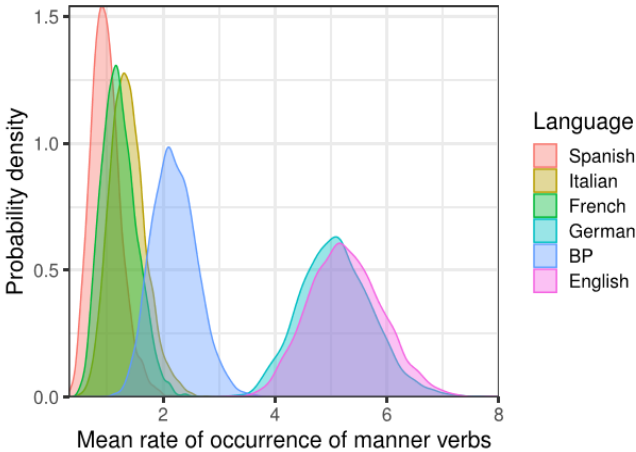
Figure 3: Posterior density for the rate of occurrence of manner verbs estimated from aggregated data.

As shown earlier, we had disaggregated data per speaker for Brazilian Portuguese and English, and we would like to take interspeaker variation within languages into account in our analysis. That is important because inferring on aggregated counts underestimates the heterogeneity between speakers and thus exaggerates differences between languages. Our decision differs from that of Hijazo-Gascón and Ibarretxe-Antuñano (6, p478), who decided to exclude atypical values from their analysis altogether.

We now sought to answer question III: whether the six languages differ concerning the mean rate of occurrence of manner verbs if we consider per-speaker variation. We analyzed all available data, including aggregated counts for Spanish, Italian, French, and German, as well as speaker-level counts for Brazilian Portuguese and English. We modified our Poisson model to

accommodate inter-speaker variation. The notation y_{ij} represents the number of manner verbs used by speaker j; $j = 1, 2, ..., j_i$ in language i. A model that includes such random effects is $y_{ij} \sim$ Poisson (λ_{ij}); $(\lambda_{ij}) = \exp(\alpha + \eta_i + \epsilon_j)$; $\alpha \sim$ Normal $(0,5)$; $\eta_i \sim$ Normal $(0,1)$; $\epsilon_j \sim$ Normal $(0,1)$.

The structure is very similar to our previous one. Still, we added a speaker-level coefficient ϵ_j to account for participant variations and used $\lambda_{ij} = \exp(\alpha + \eta_i)$ as our target quantity to make it compatible with the previous analysis. The results in Table 4 and Figure 4 demonstrate the relevance of including speaker-level data.

With within-language variation being taken into account, the mean rate of manner-verb occurrences for English drops from 5.27 to 4.35, lower than in German. That means that one or more individuals might pull the mean upwards when data are aggregated. Albeit less pronounced, the same reduction pattern (2.21 to 1.86) occurred in Brazilian Portuguese. In Table 4, we speak of the *mean* column on the left compared to the same column in Table 3 above.

Table 4: Posterior mean and 95% BCIs for the rate of occurrence of manner verbs estimated from disaggregated data

| Language	Mean	Lower	Upper
Spanish	0.98	0.51	1.58
Italian	1.37	0.80	2.08
French	1.20	0.70	1.86
Brazilian Portuguese	1.86	0.94	3.30
German	5.06	3.91	6.38
English	4.35	2.25	7.63

In addition, as expected, uncertainty in the estimates increased after we propagated uncertainty about speaker-level parameters to our language-level estimates. Results for English have much more uncertainty than those for Brazilian Portuguese, presumably due to more variability in the speaker-level data (see Figure 1). We can see that uncertainty on the graph below, with wider intervals for the Romance languages that appear as narrower curves in the graph.
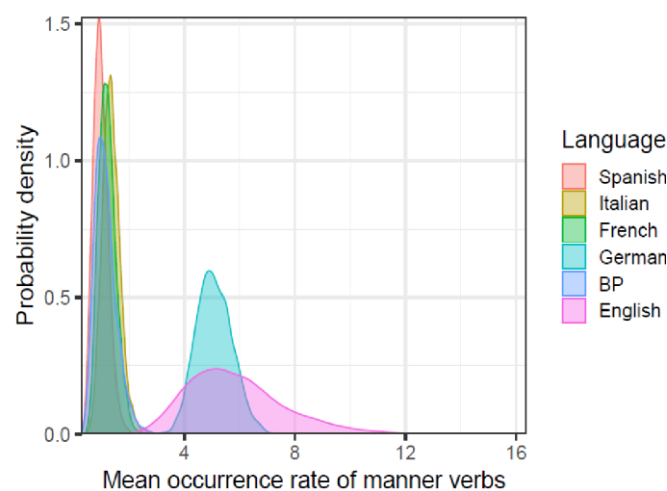
Figure 4: Posterior density for the rate of occurrence of manner verbs estimated from disaggregated data.

To complete our analysis, we plotted the individual-level rates, $\lambda_{ij} = \exp(\alpha + \eta_i + \epsilon_j)$ for both English and Brazilian Portuguese in Figure 5. These plots show that (a) there is indeed substantial within-language variation in the rate of manner-verb use and (b) most individuals are

close to the corresponding language-level mean, with only one speaker in English (speaker 3) deviating substantially from their group median. Evidence for this is that the speaker's BCI does not cover the respective language-level median the horizontal dashed line represents.
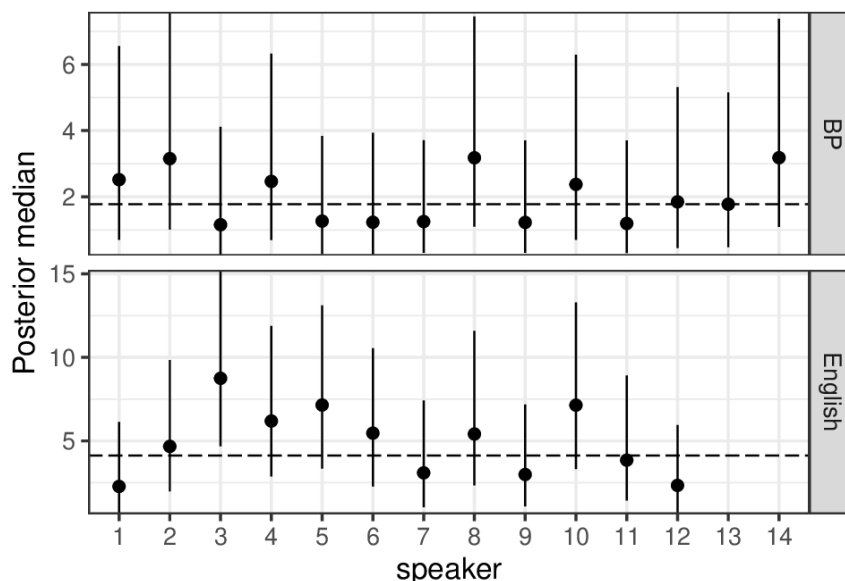
Figure 5: Posterior median and BCI for the rate of occurrence of manner verbs per speaker. Horizontal lines mark the language-level median. Please note that the speakers are not the same across BP and English.

We tested the measurement for English speaker 3, which proved not to be an outlier as strictly defined (presenting above three standard deviations from the average). Their measurement is higher than all the others, and that provides a statistically different estimate.

Overall, we were satisfied with the results obtained with the Bayesian method, and the data and code by Carvalho et al (32) is freely available. The model proved helpful, given the amount of data we used following a pattern observed in several frog-story studies. Still, for Bayesian analyses, the more observations, the better the convergence and estimates models provide.

6 Conclusions

This paper primarily aimed to include Portuguese among other Romance languages investigated in relation to Talmy's Motion Typology. According to this theory, these languages are V-framed in contrast to Germanic languages, which are S-framed. They differ because their grammars have different resources to describe motion scenes. Among these resources, one can identify a diverse inventory of motion verbs. We concentrated our efforts on comparing the use of manner verbs, which is just one of several aspects of this theory. Analyzing native speakers' data, we examined any possible intra-group (or intra-family) variation between Portuguese, Spanish, French, and Italian in this respect. Furthermore, we sought inter-group variation.

This idea originated after reading Hijazo-Gascón and Ibarretxe-Antuñano's (6) paper, which compares three other Romance languages to German. In addition to our home-collected Brazilian-Portuguese data, we included English data from the Childes Project. Due to a shared data collection protocol, we could compare these idioms using three aggregated and disaggregated data sources, a methodological issue we tackled by proposing a Bayesian statistical analysis with a few simple models. That enabled us to answer three main questions regarding the pattern of occurrence of manner verbs across these languages.

First, it became very clear that the speakers of the satellite-framed languages used a more extensive lexicon of manner verbs, which supports the commonly held idea that Germanic languages have more verbs that express this component than Romance languages. Such intergroup differences are easily obtainable in Table 1. The ratios between manner and general-motion verbs in Table 2 confirm this observation, where we also see substantial similarity between Brazilian Portuguese, Spanish, and Italian. At the same time, French had a slightly lower figure than these three. These findings echo Hijazo-Gascón and Ibarretxe-Antuñano's, according to which French speakers follow more "strictly" the V-language rhetorical style, giving little attention to manner and the path in prepositional phrases.

Our second question regarded a per-speaker rate of manner-verb occurrence within each language. The rationale here was simple: if an idiom typically encodes the manner of motion in the main predicate and has a considerably more significant number of manner verbs available for use, as S-languages do, its speakers should be able to use more of these verbs as they speak about motion events, which would manifest itself as a higher rate value. The contrary holds for V-languages. As shown in Table 3 and Figure 5, German and English fared very close in this respect, having more than twice the highest rate among the Romance languages. Given the differences discussed so far, such a result did not come as a surprise. Italian, French, and Brazilian Portuguese showed similar results, and Spanish had a slightly lower rate per speaker.

Our results confirm a typological principle that manifests across closely related languages in two individual families regarding the expression of manner. These findings pair with previous research, particularly that of Hijazo-Gascón and Ibarretxe-Antuñano, who obtained a "scarce" total number of manner verbs in all the Romance languages they investigated, with French presenting the smallest number. Additionally, the inclusion of Brazilian Portuguese and English into the analysis did not cause any substantial intra-variation in either group concerning the expression of the manner.

As for our third question, from a statistical point of view, we sought to determine the role of inter-speaker variation in the results, that is, the mean rate of occurrence of manner verbs. A few individuals might increase the language-level rate, as was the case for Brazilian Portuguese and English, for which the analysis of disaggregated data showed substantial variability between speakers in their use of motion verbs in general and manner verbs in particular. This is due to individual preferences a researcher cannot account for, such as one's rhetorical style. It especially brought the rate of manner-verb occurrences down and placed English closer to German than in the aggregated analysis.

Finally, the Bayesian analyses allowed us to contradict the assumption that data must be the same size for the six languages. We thus avoided "throwing away" data that was hard to obtain and analyzed the fourteen narratives available in our original Brazilian-Portuguese corpus rather than just twelve, as in the other languages. We could also present an alternative, perhaps more realistic, view of our findings.

7 Acknowledgments

Linguistics and for demonstrating the importance of empirical data. Our paper is supposed to embody these.

REFERENCES

1. Talmy L. Path to realization: A typology of event conflation. In: *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society* [Internet]. Berkeley; 1991 [cited 2025 Jun 08]. p. 480–519. Available from: https://journals.linguisticsociety.org/proceedings/index.php/BLS/article/view/1620/1394

2. Talmy L. *Toward a cognitive semantics: Typology and process in concept structuring*. Cambridge, MA: The MIT Press; 2000.

3. Slobin DI. The many ways to search for a frog: Linguistic typology and the expression of motion events. In: Strömqvist S, Verhoeven L, editors. *Relating events in narrative: Typological and contextual perspectives*. Vol. 2. Hillsdale: Lawrence Erlbaum Associates; 2004. p. 219–57.

4. Phelps KS. Semantic Representations of English and German Motion Verbs. *CRIL* [Internet]. 2012 Aug 1 [cited 2025 Jul 11];23. Available from: https://journals.colorado.edu/index.php/cril/article/view/311

5. Özçalişkan S, Slobin D. Learning how to search for the frog: Expression of manner of motion in English, Spanish, and Turkish. In: Greenhill A, Littlefield H, Tano C, editors. *Proceedings of the 23rd Annual Boston University Conference on Language Development.* Cascadilla Press; 1999. p. 541–52.

6. Hijazo-Gascón A, Ibarretxe-Antuñano I. Las lenguas románicas y la tipología de los eventos de movimiento. *Roman Forsch*. 2013;125(4):467–94.

7. Slobin DI. Thinking for speaking. In: *Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society*. Berkeley. 1987. p. 435–45.

8. Slobin DI. From "thought and language" to "thinking for speaking". In: Gumperz JJ, Levinson SC, editors. *Rethinking linguistic relativity*. Cambridge: Cambridge University Press; 1996. p. 70–96.

9. Filipović L, Ibarretxe-Antuñano I. Motion. In: Dabrowska E, Divjak D, editors. *Handbook of Cognitive Linguistics*. Berlin: De Gruyter Mouton; 2015. p. 527–46.

10. Mayer M. *Frog, where are you?* New York: Dial Books for Young Readers; 1969.

11. Talmy L. *Semantic structure in English and Atsugewi* [dissertation]. Berkeley: University of California at Berkeley; 1972.

12. Talmy L. Lexicalization patterns: semantic structure in lexical forms. In: Shopen T, editor. *Language typology and syntactic description*. Cambridge: Cambridge University Press; 1985. p. 57–149.

13. Slobin DI, Ibarretxe-Antuñano I, Kopecka A, Majid A. Manners of human gait: a crosslinguistic event-naming study. *Cogn Linguist*. 2014;25(4):701–41.

14. Slobin DI. Mind, code, and text. In: Bybee J, Haiman J, Thompson S, editors. *Essays on language function and language type: Dedicated to T. Givón*. Amsterdam: John Benjamins; 1997. p. 437–67.

15. Moura H, Badaracco L. Traduzindo verbos de movimento do inglês para o português: léxico e construções. *Cad Trad Florianópolis* [Internet]. 2019 May–Aug [cited 2025 Jun 10];39(2):166–83. Available from: https://periodicos.ufsc.br/index.php/traducao/article/view/2175-7968.2019v39n2p166/40002

16. Aronoff M, Fudeman K. *What is morphology?* Second edition. Oxford: Wiley-Blackwell; 2011.

17. Santos Filho DG. *A expressão do modo de movimento no português brasileiro* [dissertation]. Florianópolis: Universidade Federal de Santa Catarina; 2018.

18. Aske J. Path predicates in English and Spanish: A closer look. In: *Proceedings of the 15th Annual Meeting of the Berkeley Linguistics*. [Internet] Berkeley. 1989. [cited 2025 Jun 9] p. 1–14. Available from: http://journals.linguisticsociety.org/proceedings/index.php/BLS/article/viewFile/1753/1525.

19. Zlatev J, Yangklang P. A Third Way to Travel: The Place of Thai in Motion-Event Typology. In: Stromqvist S, Verjoven L, editors. *Relating events in narrative: typological and contextual perspectives*. Vol. 2. Mahwah, New Jersey: Lawrence Erlbaum Associates; 2004. p. 159–90.

20. Slobin DI, Hoiting N. Reference to movement in spoken and signed languages: Typological considerations. In: *Proceedings of the 20th Annual Meeting of the Berkeley Linguistics Society*. Berkeley; 1994. p. 487–505.

21. Talmy L. Properties of main verbs. *Cogn Semant*. 2016;2(2):133–63.

22. Bamberg MG. *Form and function in the construction of narratives: developmental perspectives*. [dissertation]. Berkeley: University of California at Berkeley; 1985.

23. Berman RA, Slobin DI. *Relating events in narrative: a crosslinguistic developmental study*. Hillsdale: Lawrence Erlbaum Associates; 1994.

24. Guo J, Lieven E, Budwig N, Ervin-Tripp S, Nakamura K, Özçalışkan S, editors. *Crosslinguistic Approaches to the Psychology of Language: Research in the tradition of Dan Isaac Slobin*. New York: Psychology Press; 2009.

25. Ibarretxe-Antuñano I, editor. *Motion in space across languages*. Amsterdam: John Benjamins; 2017.

26. Oliveira AA, Almeida KYB, Carmo KX, Vicente LB, Fernandes RM. A expressão da temporalidade em narrativas orais de aprendizes iniciantes de português do Brasil L2: Um estudo orientado pelo significado. *Cad Est Ling* [Internet]. 2022 [cited 2025 May 10];64:1–20,e022030. Available from: https://econtents.bc.unicamp.br/inpec/index.php/joss/article/view/20015/14519

27. Marchman VA. *Episodic structure and the linguistic encoding of events in narrative: A study of language acquisition and performance* [dissertation]. Berkeley: University of California at Berkeley; 1989.

28. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. Boca Raton: Chapman & Hall/CRC/ CRC press; 2013. Texts in Statistical Science.

29. Hijazo-Gascón A. *La expresión de eventos de movimiento y su adquisición en segundas lenguas* [dissertation]. Zaragoza: Universidad de Zaragoza; 2011.

30. Dobson, A. *An Introduction to Generalized Linear Models*. Fourth edition. Boca Raton: Chapman & Hall/CRC press; 2018. Texts in Statistical Science.

31. Gabry J, Češnovar R, Johnson A, Bronder S. *cmdstanr: R Interface to 'CmdStan'* [Computer program] R package version 0.8.1 [Internet]. 2024. Available from https://discourse.mc-stan.org, https://mc-stan.org/cmdstanr/.

32. Carvalho LM, Oliveira AA, Macario RF. *Think to Speak*. [Computer code and data]. Version A2 [Internet] Zenodo; 2025. Available from https://zenodo.org/records/15614970

APPENDIX
Computing environment

This document uses the statistical computing language R in a GNU/Linux (Ubuntu) operating system on a Dell G5 with an intel i7 processor and 16GB of RAM. See below for a complete list of the packages used.

```
## R version 4.2.2 (2022-10-31)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Pop!_OS 22.04 LTS
## Matrix products: default
## BLAS:
/usr/lib/x86_64-linux-gnu/blas/libblas.so.3.10.0
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.10.0
## ## locale:
##  [1] LC_CTYPE=en_GB.UTF-8 LC_NUMERIC=C
## [3] LC_TIME=en_GB.UTF-8 LC_COLLATE=en_GB.UTF-8
## [5] LC_MONETARY=en_GB.UTF-8 LC_MESSAGES=en_GB.UTF-8
## [7] LC_PAPER=en_GB.UTF-8 LC_NAME=C
## [9] LC_ADDRESS=CLC_TELEPHONE=C
```

```
 ## [11] LC_MEASUREMENT=en_GB.UTF-8 LC_IDENTIFICATION=C
## attached base packages:
## [1] stats  graphics grDevices utils datasets methods
## other attached packages:

## [1] kableExtra_1.3.4  lubridate_1.9.2  forcats_0.5.2
## [4] stringr_1.5.0 dplyr_1.1.2 purrr_1.0.1
## [7] readr_2.1.4 tidyr_1.3.0 tibble_3.2.1
## [10] tidyverse_1.3.2 reshape2_1.4.4 rstan_2.21.8
## [13] ggplot2_3.4.2 StanHeaders_2.21.0-7 cmdstanr_0.6.0 ##
## loaded via a namespace (and not attached):
## [1] matrixStats_1.0.0 fs_1.6.3 webshot_0.5.5
## [4] httr_1.4.4 tensorA_0.36.2 tools_4.2.2
## [7] backports_1.4.1 utf8_1.2.3 R6_2.5.1
## [10] DBI_1.1.3 colorspace_2.1-0 withr_2.5.0
## [13] tidyselect_1.2.0  gridExtra_2.3 prettyunits_1.1.1
## [16] processx_3.8.2 compiler_4.2.2 cli_3.6.1
## [19] rvest_1.0.3 xml2_1.3.3 labeling_0.4.2
## [22] bookdown_0.34  posterior_1.4.1 scales_1.2.1
## [25] checkmate_2.2.0 callr_3.7.3 systemfonts_1.0.4
## [28] digest_0.6.33 rmarkdown_2.19 svglite_2.1.1
## [31] pkgconfig_2.0.3 htmltools_0.5.4 dbplyr_2.3.0
## [34] fastmap_1.1.1 rlang_1.1.1 readxl_1.4.2
## [37] rstudioapi_0.14 farver_2.1.1 generics_0.1.3
## [40] jsonlite_1.8.7 googlesheets4_1.0.1 distributional_0.3.2
## [43] inline_0.3.19 magrittr_2.0.3 loo_2.5.1
## [46] Rcpp_1.0.11 munsell_0.5.0 fansi_1.0.4
## [49] abind_1.4-5 lifecycle_1.0.3 stringi_1.7.12
## [52] yaml_2.3.6 pkgbuild_1.4.0 plyr_1.8.8
## [55] grid_4.2.2 parallel_4.2.2 crayon_1.5.2
## [58] haven_2.5.2 hms_1.1.3 knitr_1.41
## [61] ps_1.7.5 pillar_1.9.0 codetools_0.2-18
## [64] stats4_4.2.2 reprex_2.0.2 glue_1.6.2
## [67] evaluate_0.21 data.table_1.14.8 RcppParallel_5.1.7
## [70] modelr_0.1.10 vctrs_0.6.3 Tzdb_0.4.0
## [73] cellranger_1.1.0 gtable_0.3.3 assertthat_0.2.1
## [76] xfun_0.39  broom_1.0.2 googledrive_2.0.0
## [79] viridisLite_0.4.2 gargle_1.2.1 timechange_0.2.0
```