

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Information structure drives construction selection: a quantitative investigation

Permalink

<https://escholarship.org/uc/item/2mk089nw>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Poliak, Moshe

Ricciardi, Giuseppe

Gibson, Edward

Publication Date

2026

Copyright Information

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

Peer reviewed

Information structure drives construction selection: a quantitative investigation

Moshe Poliak (moshepol@mit.edu)¹, Giuseppe Ricciardi (ricciardigiuseppe5@gmail.com)¹ &
Edward Gibson (egibson@mit.edu)¹

¹Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Abstract

In all languages, there exist multiple so-called alternations—pairs of constructions with similar truth conditions—such as the dative alternation, which consists of the double object and prepositional-phrase object constructions. Many alternations exhibit an asymmetry such that there is a canonical, general construction and a non-canonical, alternative construction with a different ordering of the verbal arguments. What is the function of such redundant constructions? We tested a proposal by Birner & Ward (2009, 1998) that the non-canonical construction is used to satisfy an information-structural constraint: placing old (given) information before new information in the sentence. In 4 experiments, in English and Italian, on the dative and locative alternations, we find experimental evidence that non-canonical constructions show preference for old-before-new information structure above and beyond canonical constructions, supporting this idea. These findings provide initial empirical evidence for a theory of how information structure may influence the use and emergence of constructions.

Keywords: language; alternations; construction grammar; discourse; information structure

Introduction

Any language in the world has many ways to express a similar message. In languages with a relatively fixed word order, like English or Mandarin, a similar message can be expressed using pairs of constructions that have largely overlapping truth conditions but different word orders. Such pairs of constructions have historically been called alternations¹ (cf. Pijpops, 2020). Consider the event in Examples 1-2, where a librarian gives a novel to a student. The event can be described using a variety of constructions that have similar truth conditions, such as the double object dative (DO) construction (1b) or the prepositional-phrase object dative (PO) construction (1a). In contrast, in languages that are considered to have a more flexible word order, like Russian or Finnish, virtually any sentence can be expressed with any ordering of the subject, verb, and object while preserving its truth conditions (subject to discourse constraints). So, whereas flexible-word-order languages have a general rule about the flexibility of word

order, fixed-word-order languages have (1) a dominant word order, used in basic, default constructions (canonical), and, (2) only in specific cases, an alternative construction with a different word order (non-canonical), whose meaning meets similar truth conditions. Why do such alternations emerge in language, and how are they used?

In the current study, we evaluate a proposal by Birner & Ward (2009, 1998) and provide experimental evidence for the idea that non-canonical constructions in alternations offer the possibility of placing old (given) material before new material—yielding an old-before-new information structure where the canonical form would produce new-before-old. More precisely, because information status is gradient rather than binary (Givón, 1983; Prince, 1981), the proposal is that non-canonical constructions allow placing **older** information before **newer** information in the sentence. On this account, the non-canonical construction is used specifically to achieve an old-before-new order. This may benefit production: speakers can begin with readily accessible material while retrieving or planning the new material, reducing pauses or disfluency that a new-first order might induce.

The proposal by Birner & Ward (2009, 1998) assumes two theoretical primitives: that an information-structural pressure for old-before-new information structure exists, and that there exist two alternatives for ordering elements: a canonical and non-canonical construction. In the rest of the introduction, we will provide a short overview of the literature motivating these theoretical primitives, illustrate the main proposal using the dative alternation in English, and finish with a preview of the empirical sections of the current manuscript.

A preference for old-before-new information structure has long been observed in English (Arnold et al., 2000; Chafe, 1976; Clark & Clark, 1977; Clark & Haviland, 1977; Clark et al., 1977; Firbas, 1966; Gundel, 1988; Halliday, 1967; Prince, 1981; Vallduví, 1990). Other languages with diverse word orders show a similar preference: for example, in Mandarin (Chen et al., 2020), Japanese (Ferreira & Yoshita, 2003), Spanish (Prat-Sala & Branigan, 2000), Arabic (Semsem & Chen, 2019), Finnish (Kaiser & Trueswell, 2004), and Czech (Šimík & Wierzba, 2017). Thus, it seems that a preference for old-before-new information structure is itself a potential cross-linguistic universal.

As for canonicity, this term is closely related to the term “markedness” (or “unmarkedness”), which is often appealed to, although without consensus on precise definitions. For

¹The term alternation has been used in the past to suggest that the two involved constructions are connected through one being derived from the other (e.g., Chomsky, 1965; Larson, 1988). When we use the term alternations in the current study we do not make such a claim (and see an argument against this view in Goldberg, 2002); instead, we simply mean that the two constructions are rough paraphrases of one another.

example, Givón (1979) proposes that markedness is associated with being structurally complex, cognitively complex, and having low frequency. Canonical word orders are the other side of the coin of markedness: they are considered basic (Greenberg, 1963) and are associated with high frequency (Hawkins, 1983), involving minimal morphological complexity (Hawkins, 1983), and being pragmatically neutral (Lambrecht, 1994). Canonicity in the sentence may also be associated with other properties of the sentence, beyond word order, such as having definite arguments, having semantically appropriate subjects and objects, and having verbs that denote actions (Siewierska, 1988).

The best-known interaction of information structure and canonicity comes from the English dative alternation (e.g., Example 1; Brown et al., 2012; Clifton and Frazier, 2004). The dative alternation consists of two constructions, PO (Examples 1a, 2a), in which the theme precedes the recipient, and DO (Examples 1b, 2b), in which the recipient precedes the theme. Although the exact definition of canonicity is not yet agreed upon in the literature, we rely on the consensus that, in the dative alternation, the PO construction is canonical and DO is non-canonical (Bresnan et al., 2007; Clifton & Frazier, 2004; Larson, 1988). (The current study also involves experiments on the locative alternation, as in Example 4, which will be described further below; we follow consensus by assuming that the word order Subject-Verb-PP is canonical and PP-Verb-Subject (the locative inversion) is non-canonical (Bresnan, 1994).

Consider Example 1, where the context introduces the theme and not the recipient, thus making PO have an old-before-new information structure and DO a new-before-old information structure. In contrast, in Example 2, the context introduces the recipient and not the theme, thus making PO have new-before-old information structure and DO old-before-new. A repeated observation in the literature is that, whereas participants do not show a preference for either information structure in PO (1a and 2a are similar, as measured by reading times or acceptability judgments, e.g., Brown et al., 2012; Clifton and Frazier, 2004), there is a substantial preference for old-before-new in DO (2b is preferable over 1b). This example shows that old-before-new may affect acceptability in only one but not the other construction in an alternation, providing existence proof for the proposal by Birner & Ward (2009, 1998).

- (1) *Context (introduces theme)*: A librarian received a novel for her birthday, but she already owned a copy of it.
 - a. The librarian gave the novel to a student.
 - b. The librarian gave a student the novel.
- (2) *Context (introduces recipient)*: A librarian was recataloguing her fiction collection with the help of a student who worked in the campus library every weekend.
 - a. The librarian gave a novel to the student.
 - b. The librarian gave the student a novel.

In the current study, we evaluate the proposal by Birner & Ward (2009, 1998) using four acceptability judgment experiments, spanning the dative and the locative constructions in English and Italian. Whereas the idea that non-canonical constructions serve old-before-new information structure may serve a pressure during production, we use acceptability judgments because of their simplicity, following a common assumption that acceptability judgments mirror usage statistics in the language (i.e., what is more likely to be produced is also rated as more acceptable; Gibson, 2025 and references within; Mahowald et al. 2016; Sprouse et al. 2013). The experiments vary by language and alternation, with each manipulating the construction (canonical or non-canonical) and the information structure of the verbal arguments that exchange positions in the alternation (old-old, old-new, new-old, new-new). The proposal that non-canonical constructions serve to place old information before new information in the sentence generates the prediction that information structure and canonicity should interact such that non-canonical constructions with new-before-old information structure should be super-additively less acceptable, as they are predicted to have the function of allowing old-before-new information structure.

Dative Experiments

In this section, we describe the English and Italian experiments that tested the interaction of canonicity and information structure in the dative alternation. In English, this experiment amounts to a replication of past findings that show that old-before-new information structure is preferable to new-before-old information structure in DO but not PO (Brown et al., 2012; Clifton & Frazier, 2004). In Italian, however, as far as we are aware, this effect has not been tested experimentally in the past. This is a crucial extension, because, in principle, it is possible that old-before-new is an idiosyncratic preference of the English DO construction, without relation to its position in the alternation, or that old-before-new preferences in the non-canonical construction is an idiosyncratic preference of the English language. Thus, the Italian dative alternation is an important case study to see if the proposal by Birner & Ward (2009, 1998) may generalize to other languages. These experiments are reported together because the Italian-language experiment used a translation of the English materials, maintaining the same design.

Method

English. All experiments followed the same design. Participants were asked to read a context sentence and a target sentence, and then rate the latter for naturalness within the context of the previous sentence on a 1-5 scale. The experimental questionnaire had 24 items. Questionnaires followed a Latin Square design, with a 2x2x2 factorial design: construction canonicity (canonical/non-canonical), the information status of the first NP (old/new), and the information status of the second NP (old/new). In the English dative, the canonical construction was PO and the non-canonical DO. An NP was

considered “old” if it was mentioned in the context sentence, in which case it was also introduced by a definite article, and “new” if it was not mentioned in the context sentence, in which case it was introduced by an indefinite article. In all items, the agent and recipient were always animate, and the theme was always inanimate. See Examples 1-2 above for sample materials (these are the 4 conditions that have either old-new or new-old information structure). The experiment was conducted online using the MTurk platform, where we recruited 63 native English speakers. **Italian.** The Italian materials were translated from English. The key difference is that, unlike in English, where PO has a preposition before the recipient and DO does not, in Italian the recipient retains the article across all order variations: either NP-PP (theme-recipient, canonical) or PP-NP (recipient-theme, non-canonical; see Example 3). The experiment was conducted online using the Prolific platform, where 50 native speakers of Italian were recruited.

(3) **The Italian dative alternation.**

a. *canonical (NP-PP)*

La bibliotecaria diede un
DEF.F.SG librarian give.PST.3SG INDF.M.SG
romanzo alla studentessa.
novel to=DEF.F.SG student.F.SG

‘The librarian gave a novel to the student.’

b. *non-canonical (PP-NP)*

La bibliotecaria diede alla
DEF.F.SG librarian give.PST.3SG to=DEF.F.SG
studentessa un romanzo.
student.F.SG INDF.M.SG novel

‘The librarian gave a novel to the student.’

Results

See Figure 1 for a visualization of the data. The data were analyzed and visualized in R using the tidyverse packages (Wickham et al., 2019) and formal analysis was conducted using the brms package (Bürkner, 2017). For simplicity, we only present results from the old-new and new-old conditions, sum coded (old-new = -0.5, new-old = 0.5). Canonicity was also sum coded (-0.5 = canonical, 0.5 = non-canonical). As for the other information structure conditions, it was always the case that old-old was the most acceptable information structure, and new-new was either the least acceptable, or only more acceptable than new-old. For each experiment, we regressed acceptability on the information-structure and canonicity factors and their interaction using an ordinal Bayesian regression and the default priors from the brms package (flat priors over fixed effects). We used the maximal random effects structure justified by the design (Barr et al., 2013), which involved random intercepts for participants and items, and random slopes that mirrored the fixed effects within participants and items. All models used 4000 iterations (2000 warmup), completed without divergent transitions, with all parameters having R-hat values below 1.01. **English.** On average, the non-canonical double object construction was less acceptable

than the prepositional object (Posterior Mean (PM) = -1.84; 95% Credible Interval (CrI) = [-2.67, -1.06]); on average, new-old information structure was less acceptable than old-new (PM = -0.94, CrI = [-1.56, -0.38]); and, critically, there was substantial evidence for an interaction such that old-new information structure was more preferable to new-old in the double object construction than in the prepositional object construction (PM = -2.93, CrI = [-4.43, -1.54]). **Italian.** As in English, on average, the non-canonical PP-NP was less acceptable than the canonical NP-PP (Posterior Mean (PM) = -1.98; CrI = [-2.69, -1.05]). Unlike in English, old-new information structure was not on average substantially different from new-old (PM = -0.02, CrI = [-0.58, 0.55]). Critically, like in English, there was substantial evidence for an interaction such that new-old information was more strongly dispreferred in the non-canonical construction relative to the canonical construction (PM = -1.69, CrI = [-2.71, -0.74]).

Locative Experiments

The dative experiments found an interaction between canonicity and information structure, thus supporting the proposal that the non-canonical construction may be used to achieve an old-before-new information structure. The locative alternation provides an important additional test of the same idea, as **it is possible that old-before-new preferences in non-canonical (recipient-theme) dative constructions are specific to the semantics of dative events.** For example, perhaps the non-canonical dative construction did not evolve to satisfy information structure per se; instead, a pressure to put specifically recipients, which are often animate and salient in context, early in the sentence may have resulted in the non-canonical dative construction having specific information structure preferences. For other constructions, where there is no such semantic regularity, the non-canonical construction may be unaffected by information structure. Thus, the dative alternation allows us to test whether old-before-new preferences are associated with the non-canonical construction more broadly, or only within a specific alternation.

Method

English. The design was identical to the English dative, with the difference that the canonical construction was NP-V-Loc and the non-canonical construction was the locative inversion, Loc-V-NP (Example 4). The experiment was conducted online using the MTurk platform, where we recruited 55 native English speakers. **Italian.** The Italian materials were again translated from English. (Example 4). The experiment was conducted online using the Prolific platform, where 50 native Italian speakers were recruited.

(4) **The locative alternation (English and Italian).**

a. *canonical (NP-V-Loc)*

English: The weapon lay behind the box.

Italian:

L' arma giaceva dietro
DEF.F.SG weapon lie-IPFV.PST.3SG behind

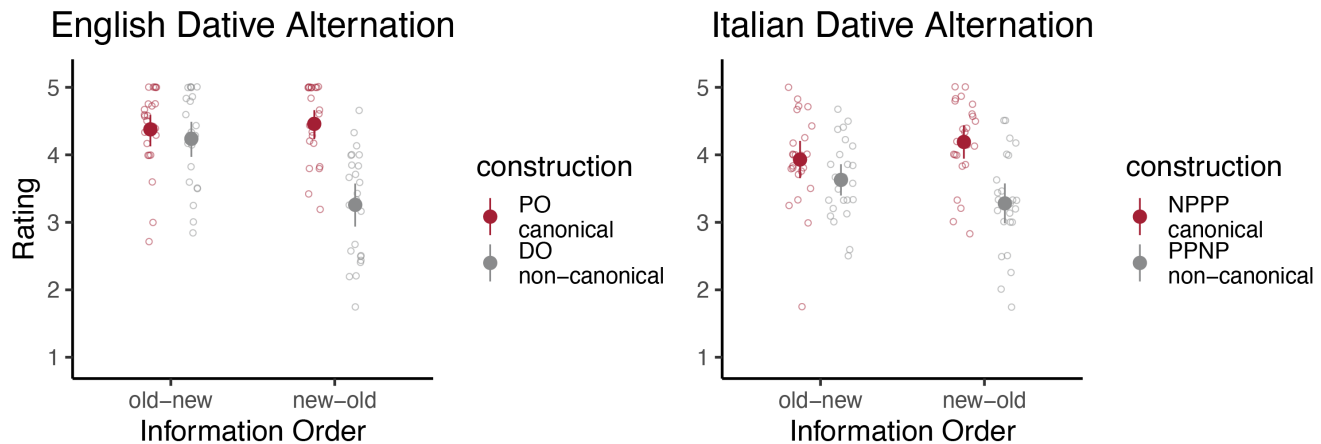


Figure 1: Responses in the dative experiments. Large points represent overall means and error bars represent 95% Bootstrapped Confidence Intervals over item means. Hollow circles represent individual item means. Note that the main effects (i.e., baseline differences in acceptability) are not comparable across languages.

- alla scatola.
to=DEF.F.SG box
'The weapon lay behind the box.'
- b. *non-canonical (Loc-V-NP)*
English: Behind the box lay the weapon.
Italian:
Dietro alla scatola giaceva l'
behind to=DEF.F.SG box lie-IPFV.PST.3SG DEF.F.SG
arma.
weapon
'Behind the box lay the weapon.'

Results

See Figure 2 for a visualization of the data. All model specifications were identical to the dative experiments. **English.** On average, the non-canonical Loc-V-NP was less acceptable than the canonical NP-V-Loc (PM = -1.61, CrI = [-2.39, -0.86]); on average, new-old information structure was less acceptable than old-new (PM = -0.36, CrI = [-0.74, 0]), with no substantial evidence for an interaction, although the estimate pointed in the predicted direction (PM = -0.33, CrI = [-1.04, 0.40]). **Italian.** On average, the non-canonical Loc-V-NP construction was less acceptable than the canonical NP-V-Loc (PM = -1.66, CrI = [-2.30, -1.04]) and new-before-old information structure was less acceptable than old-before-new (PM = -1.64, CrI = [-2.19, -1.12]). Critically, unlike English, the model quantified substantial evidence that the non-canonical construction dispreferred new-old information structure more than the canonical construction (PM = -1.90, CrI = [-2.89, -0.90]).

Interim Discussion

Both dative experiments, in English and in Italian, found an interaction between canonicity and information structure that supports the idea that non-canonical constructions are used to achieve old-before-new information structure (Birner & Ward,

2009; Birner & Ward, 1998). Generally, the locative experiments show a similar pattern, except that the evidence for the interaction was not substantial in English. Thus, it seems that old-before-new preferences in the non-canonical construction are not merely a property of the dative alternation. However, this raises the question of why the English locative experiment did not produce conclusive evidence for an interaction between canonicity and information structure. It could be the case that the proposal by Birner & Ward (2009, 1998) simply does not apply to all languages and alternations, but rather to some subset that has not yet been identified. However, an alternative explanation is that the English locative alternation is different in that the non-canonical construction, the locative inversion, is uniquely rare. If it is so rare as not to be available to comprehenders, it is not a true alternant with the canonical construction. If this is true, rare non-canonical constructions should not show a preference for old-before-new information structure, as they do not provide an alternative to the canonical construction in practice. To quantitatively evaluate the frequency of the English locative inversion relative to the other constructions, we conducted a corpus analysis, reported below.

Corpus Analysis

The 4 experiments reported here use 2 pairs of constructions (dative, locative) in both English and Italian. We conducted a corpus analysis to find the frequencies of these constructions. In English, we used the Corpus of Contemporary American English (CoCA; Davies, 2008), and in Italian we used the Corpus of Written Italian (CORIS; Rossini Favretti et al., 2002; Tamburini, 2022). We extracted all the appearances of any of the forms of the verbs that were used in the experiments using regular expressions from the corpus. Subsequently, we sampled 10,000 sentences per verb (or the entire set, for less-frequent verbs) and dependency-parsed them using Stanza, a

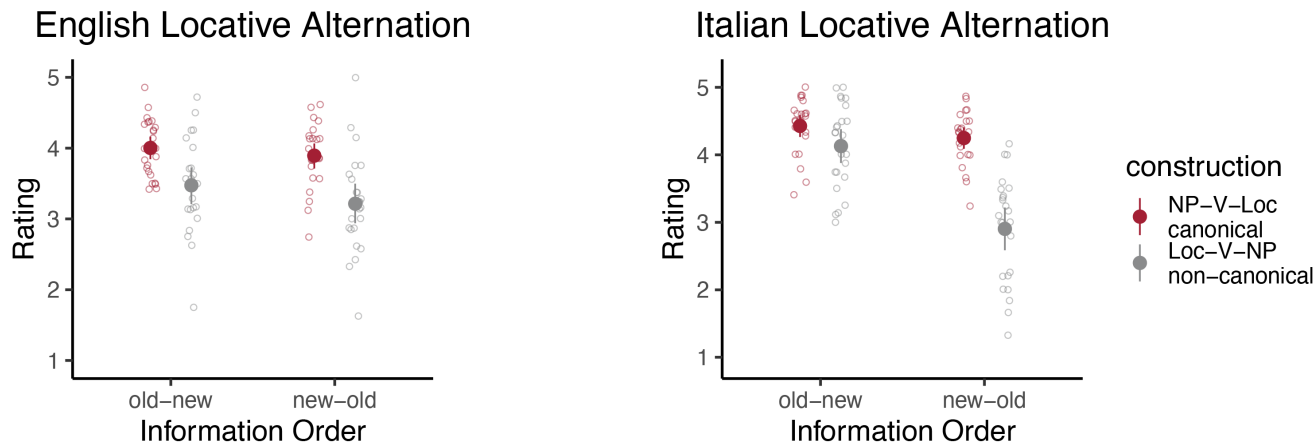


Figure 2: Responses in the locative experiments. Large points represent overall means and error bars represent 95% Bootstrapped Confidence Intervals over item means. Hollow circles represent individual item means. Note that the main effects (i.e., baseline differences in acceptability) are not comparable across languages.

neural-net dependency parser (Qi et al., 2020).

For the dative constructions, we extracted all the right-dependencies of the verb and examined their dependency relations. For English, if the verb had an indirect object and a direct object after it, the sentence was classified as DO. If the verb had a direct object, and, to the right of it, an oblique with the appropriate dative preposition for the verb (to/for), the sentence was classified as PO. For Italian, we identified sentences where the verb had 2 right-dependencies, one a direct object and the other an object with a dependent “a” preposition to the left of it (parallel to English “to”), proceeding to classify them as NP-PP and PP-NP. The locative analysis was similar to the dative. We filtered sentences to contain any of the verbs and locative prepositions from the experiments. Sentences were categorized as NP-V-Loc if the subject appeared before the verb, and the locative (a noun with one of the prepositions used in the experiment as its dependent) after the verb. Sentences were categorized as Loc-V-NP if the subject appeared after the verb and the locative before it.

Results (see Table 1). Dative. In English, the most frequent construction was the non-canonical DO (76 appearances out of 133 identified datives), whereas the most frequent construction in Italian was the canonical NP-PP (32 appearances out of 43 identified datives). **Locative.** In both languages, the most frequent construction was the canonical NP-V-Loc (English: 209 out of 212 locatives; Italian: 115 out of 174 locatives).

These results reveal that the English non-canonical locative inversion (Loc-V-NP) is substantially less frequent than the other constructions, both in terms of its raw frequency (3 observations of all sampled sentences) and in terms of the proportion of locatives that appeared in this construction (1.4%). This supports the explanation that the English locative experiment did not yield substantial evidence in favor of an interaction between information structure and canonicity due

to the frequency of the non-canonical construction: the English non-canonical locative Loc-V-NP is not frequent enough to be available as an alternative to the canonical NP-V-Loc.

Table 1: Results of the corpus analysis on the constructions in the English and Italian experiments. The counts are from the analyzed sample (up to 10,000 sentences per verb).

Construction	English	Italian
Dative, canonical	57/133 (42.9%)	32/43 (74.4%)
Dative, non-canonical	76/133 (57.1%)	11/43 (25.6%)
Locative, canonical	209/212 (98.6%)	115/174 (66.1%)
Locative, non-canonical	3/212 (1.4%)	59/174 (33.9%)

Discussion

The current study sought to experimentally test a proposal by Birner & Ward (2009, 1998), according to which one key use of non-canonical constructions in alternations is to allow positioning old information before new information in the sentence. This proposal generates the prediction that the non-canonical construction should have superadditively low acceptability if it appears with a new-before-old information structure, as in this case, the use of the non-canonical construction would be unjustified.

We conducted four experiments in English and Italian, using the dative and locative constructions. The experiment in English on the dative alternation replicated an interaction that was found in previous research (Brown et al., 2012; Clifton & Frazier, 2004), wherein new-before-old information structure is more dispreferred in the non-canonical DO than in the canonical PO, which aligns with the proposal in Birner & Ward (2009, 1998). We sought to investigate whether the same effect might be more general, extending to the dative alternation in another language. Thus, we conducted the Italian

experiment, which used translated materials from the English experiment, and found a similar data pattern: new-before-old is more strongly dispreferred in the non-canonical construction (PP-NP) than in the canonical construction (NP-PP). This finding suggests that the non-canonical construction in the dative alternation shows preference for old-before-new information structure in at least 2 languages, and merits future investigation across a broader range of languages.

Subsequently, we turned to a different alternation in these 2 languages: the locative alternation. Unlike the dative alternation, which involves different orders of the post-verbal arguments, the locative alternation in English and Italian involves different orders of the subject (canonically pre-verbal) and the locative (canonically post-verbal). In the case of the locative alternation, we found mixed results: the Italian experiment, as before, yielded a superadditive preference for old-before-new in the non-canonical construction; the English experiment, however, only yielded anecdotal evidence in favor of this effect. We hypothesized that perhaps the non-canonical locative is not available to comprehenders in English, because it is too rare. We conducted a corpus analysis to investigate this option and found that, indeed, the non-canonical locative in English (the locative inversion) is an order of magnitude less frequent than any of the other constructions in the current study. This suggests a possible limitation to the proposal by Birner & Ward (2009, 1998): the non-canonical construction is used for allowing old-before-new information structure *when it is not too rare*.

How rare a construction needs to be in order to qualify as “too rare to be in an alternation” remains an open question. In the current study, we reported a corpus analysis that found that the locative inversion accounts for 1.4% of the appearances of the locative alternation in English. The second rarest non-canonical alternation is the Italian non-canonical dative, which accounts for 25.6% of the appearances of the dative alternation in Italian. Thus, based on the current study, it is plausible that “too rare” is a line between these 2 estimates (1.4%-25.6%). However, we acknowledge that we cannot truly know this given the existing data, and further work needs to verify whether the cognitive representations of what is called “alternation” in the literature vary based on the frequency of the non-canonical alternant.

These results align with findings from Finnish, a flexible-word-order language with a canonical SVO word order. In their self-paced reading experiment, Kaiser and Trueswell (2004) found that an old-before-new information structure reduced the reading time cost of OVS (non-canonical) sentences; in their eye-tracking experiment, they found that participants were more likely to make anticipatory eye movements to a new character on the screen before hearing the last nominal element in OVS sentences, relative to SVO. They interpret these findings to mean that comprehenders expect the non-canonical word order OVS to appear with old-before-new information structure, which aligns with the broader proposal by Birner & Ward (2009, 1998). In contrast, Fedorenko and Levy (2007)

conducted a similar study using self-paced reading in Russian, another flexible-word-order language with a canonical SVO word order, and found no indication of an effect of information structure on the reading times of OVS sentences. Thus, it is not necessarily the case that all flexible-word-order languages use non-canonical word orders to achieve old-before-new information structure (Evans & Levinson, 2009). Alternatively, as self-paced reading is known to be noisy, perhaps a more reliable approach, such as acceptability ratings, may have yielded results that are more aligned with the Finnish data (Gibson, 2025).

The current work, however, concerns languages with more rigid word orders: English and Italian. We investigated the notion that perhaps in these kinds of languages, which use pairs of alternating constructions rather than a general syntactic flexibility, the non-canonical construction is used to achieve old-before-new information structure. That is, perhaps languages with a more rigid word order develop non-canonical variants in contexts where the pressure to front old information is the greatest. For example, in dative events, recipients are often animate and salient in the context, whereas the theme is often inanimate; this may make recipients likely to be accessible early in production, motivating their positioning before the theme in the sentence, resulting in the non-canonical DO construction and the dative alternation.

Construction selection is probabilistic and is also affected by factors other than information structure, such as animacy and nominal form (Bresnan et al., 2007). We tried to hold such external factors constant in the experiments above, by using only common nouns, and, in the dative alternation, using exclusively animate subjects and recipients and inanimate themes. This allowed us to investigate the effects of information structure on construction selection, finding that, across the locative and dative alternations, in English and Italian, the non-canonical form was associated with old-before-new information structure. This evidence converges with the proposal by Birner & Ward (2009, 1998), a strength of which is its generality: in any human language, some elements in discourse are old and some are new. Thus, the idea that non-canonical constructions serve placing old information before new may recur across constructions and languages, and may even emerge as a cross-linguistic generalization.

References

- Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. Newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1), 28–55.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Birner, B. J., & Ward, G. (2009). Information structure and syntactic structure. *Language and Linguistics Compass*, 3(4), 1167–1187.

- Birner, B. J., & Ward, G. L. (1998). *Information status and noncanonical word order in English*. John Benjamins Publishing.
- Bresnan, J. (1994). Locative inversion and the architecture of universal grammar. *Language*, 72–131.
- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In *Cognitive foundations of interpretation* (pp. 69–94). KNAW.
- Brown, M., Savova, V., & Gibson, E. (2012). Syntax encodes information structure: Evidence from on-line reading comprehension. *Journal of Memory and Language*, 66(1), 194–209. <https://doi.org/10.1016/j.jml.2011.08.006>
- Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic*.
- Chen, J., Narasimhan, B., Chan, A., Yang, W., & Yang, S. (2020). Information structure and word order preference in child and adult speech of Mandarin Chinese. *Languages*, 5(2), 14.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.
- Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. Harcourt Brace Jovanovich.
- Clark, H. H., Haviland, S., & Freedle, R. O. (1977). *Discourse production and comprehension*. Ablex.
- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Ablex Publishing Corporation.
- Clifton, C., & Frazier, L. (2004). Should given information come before new? yes and no. *Memory & Cognition*, 32(6), 886–895. <https://doi.org/10.3758/bf03196867>
- Davies, M. (2008). The corpus of contemporary American English (COCA): 560 million words, 1990-present [(2008–)]. <https://www.english-corpora.org/coca/>
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32(5), 429–448.
- Fedorenko, E., & Levy, R. (2007). Information structure and word order in Russian sentence comprehension. *Poster presented at the 20th CUNY conference on human sentence processing, La Jolla, CA*.
- Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research*, 32(6), 669–692.
- Firbas, J. (1966). Non-thematic subjects in contemporary English. *Travaux Linguistiques de Prague*, 2(1), 239–256.
- Gibson, E. (2025). *Syntax: A cognitive approach*. MIT Press.
- Givón, T. (1979). *On understanding grammar*. Academic Press.
- Givón, T. (Ed.). (1983). *Topic continuity in discourse: A quantitative cross-language study*. John Benjamins.
- Goldberg, A. E. (2002). Surface generalizations: An alternative to alternations.
- Greenberg, J. H. (Ed.). (1963). *Universals of language*. MIT Press.
- Gundel, J. K. (1988). Universals of topic-comment structure. In M. Hammond (Ed.), *Studies in syntactic typology* (pp. 209–242). John Benjamins B. V.
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, 3(2), 199–244.
- Hawkins, J. A. (1983). *Word order universals*. Academic Press.
- Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94(2), 113–147. <https://doi.org/10.1016/j.cognition.2004.01.002>
- Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents* (Vol. 71). Cambridge University Press.
- Larson, R. K. (1988). On the double object construction. *Linguistic Inquiry*, 19(3), 335–391.
- Mahowald, K., Graff, P., Hartman, J., & Gibson, E. (2016). Snap judgments: A small N acceptability paradigm (SNAP) for linguistic acceptability judgments. *Language*, 92(3), 619–635. <https://doi.org/10.1353/lan.2016.0052>
- Pijpops, D. (2020). What is an alternation? six answers. *Belgian Journal of Linguistics*, 34(1), 283–294. <https://doi.org/10.1075/bjl.00053.pij>
- Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, 42(2), 168–182. <https://doi.org/10.1006/jmla.1999.2668>
- Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255).
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages.
- Rossini Favretti, R., Tamburini, F., & De Santis, C. (2002). CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In A. Wilson, P. Rayson, & T. McEnery (Eds.), *A rainbow of corpora: Corpus linguistics and the languages of the world* (pp. 27–38). Lincom Europa.
- Semsem, M., & Chen, J. (2019). The use of word order to mark information status in adult and child Arabic. *Proceedings of the 30th Western Conference on Linguistics*, 173–178.
- Siewierska, A. (1988). *Word order rules*. Croom Helm.
- Šimík, R., & Wierzbica, M. (2017). Expression of information structure in West Slavic: Modeling the impact of prosodic and word-order factors. *Language*, 93(3), 671–709.
- Sprouse, J., Schütze, C. T., & Almeida, D. (2013). A comparison of informal and formal acceptability judgments using

- a random sample from Linguistic Inquiry 2001–2010. *Lingua*, 134, 219–248.
- Tamburini, F. (2022). I corpora del FICLIT, università di Bologna: CORIS/CODIS, BoLC e DiaCORIS. In E. Cresti & M. Moneglia (Eds.), *Corpora e studi linguistici. atti del LIV congresso internazionale di studi della società di linguistica italiana, online, 8–10 september 2021* (pp. 189–197). SLI. <https://doi.org/10.17469/O2106SLI000012>
- Vallduví, E. (1990). *The informational component* [Doctoral dissertation, University of Pennsylvania].
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>