

## Evidence in children’s input for the meanings of 3-argument constructional homonyms

In English, a double-object ditransitive entails intended causation of possession (e.g. (1)). In these sentences, Joyce has the thematic role of Goal (also called Recipient).

- 1a. Andrew threw Joyce the ball.
- 1b. Andrew baked Joyce a cake.

But in some languages, like Spanish and French, the same *surface morphosyntactic pattern* used for double-object ditransitives is constructionally homonymous [1] with surface patterns used to express other meanings. Some of these meanings are illustrated in the Spanish examples in (2). In 2a, Sara has the thematic role of Goal, but in 2b she has the thematic role of Affectee, and in 2c Possessor (e.g. [2]).

- 2a. Adela le dio el libro (a Sara).  
Adela 3SG.DAT gave the book (DAT Sara)  
“Adela gave Sara the book.”
- 2b. Adela le rompió el radio (a Sara).  
Adela 3SG.DAT broke the radio (DAT Sara)  
“Adela broke the radio on Sara.”
- 2c. Adela le vio el pelo (a Sara).  
Adela 3SG.DAT saw the hair (DAT Sara)  
“Adela saw Sara’s hair.”

The acquisitional puzzle that constructional homonyms, as [3] explains, is as follows. To learn the abstract differences between constructions, there need to be cues in the input that signal those differences. But the defining nature of a set of constructional homonyms is that no such cue exists in the surface patterns of these sentences.

We present evidence from two studies - (1) a clustering analysis of language model embeddings of French three-argument constructional homonyms and (2) a corpus analysis of child-directed speech in three languages (English, French, and Spanish) - to argue that successful learning depends on the learner’s having certain expectations about the possible meanings and structures of three-argument sentences.

### Study 1: LLMs do not have human-like representations of constructional meanings

*Method.* We extracted parents’ utterances from all of the French-language corpora in CHILDES [4] that matched the surface pattern **Pronoun.DAT (Aux) Verb Determiner Noun**. In this pattern, the dative pronoun can refer to a Goal (as in 3a) or an Affectee (as in 3b). Two examples of extracted sentences are given in (3). A native speaker of French hand-coded the utterances to mark those that unambiguously refer to either a Goal or Affectee.

- 3a. je vais te donner une bouteille !  
I go 2SG.DAT give a bottle  
“I’m going to give you a bottle!”
- 3b. tu lui sèches les cheveux ?  
you 3SG.DAT dry the hair  
“Are you drying his hair?”

We obtained embeddings from a French-language BERT-type model called ModernCamembert [5] using the HuggingFace Transformers library [6]. Other language models were also tested, but findings were similar. We ran KMeans clustering on dimensionally-reduced embeddings.

**Results.** We report pairwise precision and recall results for various values of  $k$  (number of clusters) in Table 1. This metric evaluates whether the clusters of embeddings align with the hand-coded (gold) speaker labels. Higher precision reflects a higher ratio of true positives to false positives (i.e., a single cluster does not mix different constructional meanings), and higher recall reflects a higher ratio of true positives to false negatives (i.e., different constructional meanings are not spread across clusters).

number of clusters (k)	pairwise precision	pairwise recall
2	.598	.861
3	.599	.439
4	.599	.437
10	.609	.114
20	.604	.068

Table 1: Pairwise precision and recall values for clusters of ModernCamembert embeddings

### Study 2: Informative cues exist in children’s input to differentiate English from French/Spanish constructional homonymity, given expectations about distribution

**Method.** We follow a corpus analytic and theoretical procedure, laid out by e.g. [3], [7], of identifying an *informative indirect cue* from which a child could learn the mappings between surface strings and abstract structure (in this case, meaning). Say a learner expects that some three-argument clauses are non-homonymous ditransitives (i.e. can express Goals only, as in English) and that others are homonymous with Affectee and/or Possessor constructions (as in Spanish and French). Our hypothesis was that she can track the distribution of *non-transferable objects*, e.g. body parts, in certain surface patterns, to figure out if the pattern is non-homonymous (English ditransitives) vs. homonymous (Spanish/French ditransitives). If the pattern shows a significant reduction in non-transferable objects compared to the rest of the language, that could be a signal that objects used with the pattern must be transferable, and hence that the pattern is restricted to expressing the Goal constructional meaning (as in English).

**Results.** We analyzed 1,222,345 English, 314,086 French, and 153,648 Spanish sentences matching the patterns of the sentences shown in the examples above, again from CHILDES [4], and annotated them based on the presence of a body-part noun. We compared the rate of body-part nouns in the pattern to body-part nouns outside of the pattern in that language. Table 2 shows results in terms of odds ratios (Fisher’s exact test,  $p=.05$ ). An odds ratio less than 1 indicates that there are fewer body-part nouns in the pattern than outside of the pattern. English shows such a reduction, consistent with its non-homonymous pattern.

Language	Odds ratio	p (OR < 1)
English	.48	$7.06e^{-27}$
French	2.38	1.0
Spanish	3.33	1.0

Table 2: Odds ratios for the count of body parts in the pattern compared to outside the pattern  
Discussion. Taken together, the results of these studies suggest that the difference in constructional meaning is not represented by LLMs but that cues for learning it are available in the input, if the learner knows what to look for and how to use it.

References. 1.Chomsky 1957, *Syntactic Structures*. 2.Demonte 1995, *Dative alternation in Spanish*. 3.Becker 2014, *The acquisition of syntactic structure*. 4.MacWhinney 2000, *CHILDES*. 5.Antoun et al. 2025, ModernBERT. 6. Wolf et al. 2019, HuggingFace’s Transformers. 7.Viau&Lidz 2011, *Selective learning in the acquisition of Kannada ditransitives*.