

Jennifer Arnold

Re-mention frequency is higher for goals than sources in transfer events

Technical Report #6, UNC

To cite:

Arnold, J. (2021). Re-mention frequency is higher for goals than sources in transfer events. Technical Report #6. UNC Language Processing Lab, Department of Psychology & Neuroscience, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina.

We examined how speakers naturally use transfer verbs, for the purpose of understanding whether some patterns of reference are more frequent, and thus more predictable, and how this predictability relates to topicality measures. The database included utterances with both goal and source arguments¹. For both datasets, we analyzed naturally-occurring speech from the Fisher Corpus (Cieri, Graff, Kimball, Miller & Walker, 2004, 2005), which is a collection of telephone conversations. Similar to the Switchboard corpus, participants in this project were asked to speak about randomly generated topics from a list.

This analysis was directed at two questions. First, are some thematic roles more likely than others to be mentioned again? That is, what is the frequency of re-mention for different roles? We examined this by counting the frequency with which each thematic role was mentioned in the immediately following clause. For example, in (1a) both the source [they] and the goal [me] are referred to in the next clause, while in (1b), the source [the Romanian army] is referred to but not the goal.

(1a) They emailed it to me and I never checked that email.

¹ Typically transfer verbs also include a theme argument, i.e. the thing that is transferred, but our focus here is on the behavior of the goal and source roles.

- (1b) They [the Americans] bought them [uniforms] second hand from the ur- the Romanian army. You know who sold them to them.

Our second purpose was to understand how referential frequency relates to discourse topicality. Some theories view predictability as a property of topicality (e.g., Givón, 1983; Prince, 1981) and some do not (Rohde & Kehler, 2014). This raises questions about whether predictable thematic roles (goals and implicit causes) tend to have other topical properties in natural discourse. To assess this, we examined the frequency with which different thematic roles co-occurred with other indicators of topicality, namely givenness, pronominalization, prominence on a person hierarchy (1st/2nd vs. 3rd), and animacy.

Selection criteria

We searched for the transfer and emotion verbs listed in Table 1, including all files in the Fisher Corpus. We began with the files in part 2 (files 058-116), and extended our search to part 1 (files 001-057) if we did not find enough tokens. We limited the search to verbs in the past tense, using a sequence of grep searches, searching for transfer verb followed by the preposition (e.g., *bought...from, emailed...to*). These tokens were examined for the criteria listed in Table 2, with the goal of identifying between 10-25 tokens for each verb. Based on our first (rough) examination, verbs with fewer than 10 usable tokens in the sample were excluded from the sample. However, some tokens were later discarded due to ambiguity or extreme disfluency, leaving fewer than 10 tokens for some verbs. In addition, we found that our initial pass identified too few stimulus-experiencer tokens, so we later included two verbs that had 5-7 usable tokens (*disappointed, impressed*).

Table 1. Transfer verbs

Verb type	Verbs	Included in analysis (n for each verb)
Goal-source	accepted from, borrowed from, grab from, bought from, catch from, got from, inherit from, learned from, received from, rented from, snatched from, took from	Bought (12), got (20), learned (24), took (24)
Source-goal	emailed to, fax to, gave to, handed to, hurled to, loaned to, lobbed to, mailed to, offered to, passed to, pay to, rent-to to, sent to, showed to, taught to, told to, throw to, toss to	Emailed (6), gave (23) handed (9), passed (22), sent (27)

The criteria used for inclusion of tokens are shown in Table 2. The logic behind these criteria follows from our goal to understand the frequency with which speakers refer to entities appearing in particular thematic roles. Thus, we selected only clauses that contained both critical entities for each verb type, controlling for the syntactic structure of the clause, and only in cases where there was a following utterance in which a reference could be made.

Table 2. Criteria for inclusion in analysis

Criterion	Example
The verb must be used with both critical roles (goal and source for transfer verbs; experiencer and stimulus for emotion verbs); transfer verbs must occur with “to” or “from”	<u>Not usable</u> : I even uh bought one of the entertainment books through my daughter's school so I use those coupons from that <u>Usable</u> : they just kind of like bought one on the internet from somebody else
The two critical roles must be entities, either animate or inanimate, but not clauses	<u>Not usable</u> : I learned it from watching this or that <u>Usable</u> : yeah only once I got someone from Arizona
The source-goal transfer verbs must occur in prepositional form, not double object form	<u>Not usable</u> : if you gave her some of the million she'd be a friend to you <u>Usable</u> : like if they gave it to you
The clause must be a tensed main or subordinate clause. Relative clauses and infinitival clauses not okay.	<u>Not usable</u> : that's the site that I learned about this from <u>Not usable</u> : the people that we bought the house from they actually won the west alice beautification award for how the yard was set up

<p>Disfluency, unfinished fragments and discourse markers are excluded for the purpose of clause inclusion; excluded material shown here in brackets</p>	<p><u>Usable</u>: if I got a phone call from the principal she passed it on to me {and I haven't you know} I've only read a part of it we sent him {to um} to private school instead and then he went to public college we sent {um you know} a lot of food and stuff over to Iraq and Afghanistan and they were still against America that kind of scared me because {I don't know it's} what's the UN for</p>
--	--

Statistical analysis

All statistical analyses used SAS proc glimmix to perform a logistic regression, given that our dependent measures are binary (e.g., is the subject mentioned in the next sentence or not; is the subject animate or not, etc.) We centered all predictors that were involved in interactions (i.e., verbtype and coherence relation). The verb was the random intercept in all models, except where the model estimated it to be zero, in which case it was excluded. Because of the small number of verbs in our analysis, we did not include random slopes. For each question, we examined the frequency of mentioning the subject and nonsubject entities in separate models.

Reference continuation

For each token in the database, we coded whether the speaker referred back to the critical entities (goal, source, experiencer, stimulus). Our analysis focuses on the broadest definition of continued reference, including both “direct” and “indirect” reference. In direct reference, the following clause refers directly and entirely to an entity in the target clause. For example, in (1), both the experiencer “we” and the stimulus “our freedom” are mentioned directly in the following

clause. Note that direct reference does not have to use the same words, as pronouns are frequently used to refer to recently mentioned entities, like “it” in (2).

(2) **Target clause:** we appreciated our freedom

Following clause: and we liked it

Indirect reference includes cases where the following clause refers to either a superset or subset of the critical referent. For example, consider “I hated Melissa / and Mo Jo we hated Mo Jo too.” The experiencer in the target clause is “I”, and the following clause mentions “we”, which is a superset of “I”. Other examples of indirect reference include “her kids”... “she”; “I”... “the girl I got matched with last time”; “him”... “his father”. There were very few items with indirect reference (4% of the subject and 5% of the objPP references), so we selected the broader category for analysis, including all cases of either direct or indirect reference.

The first question is whether some entities are more likely to be mentioned in the immediately following clause. Figure 1 shows that goals tend to be mentioned more often than sources, for almost all verb types. We assessed the reliability of this pattern in two models, one where subject continuation was the dependent measure, and one where object of preposition continuation was the separate measure. For subjects, *verdtype* was a marginally significant predictor of continuation ($\beta = 0.655 (0.31)$, $t = 2.1$, $p = .07$). For nonsubjects, *verdtype* was a significant predictor ($\beta = -0.79 (.31)$, $t = -2.53$, $p = .01$; the random intercept was estimated to be zero in this model).

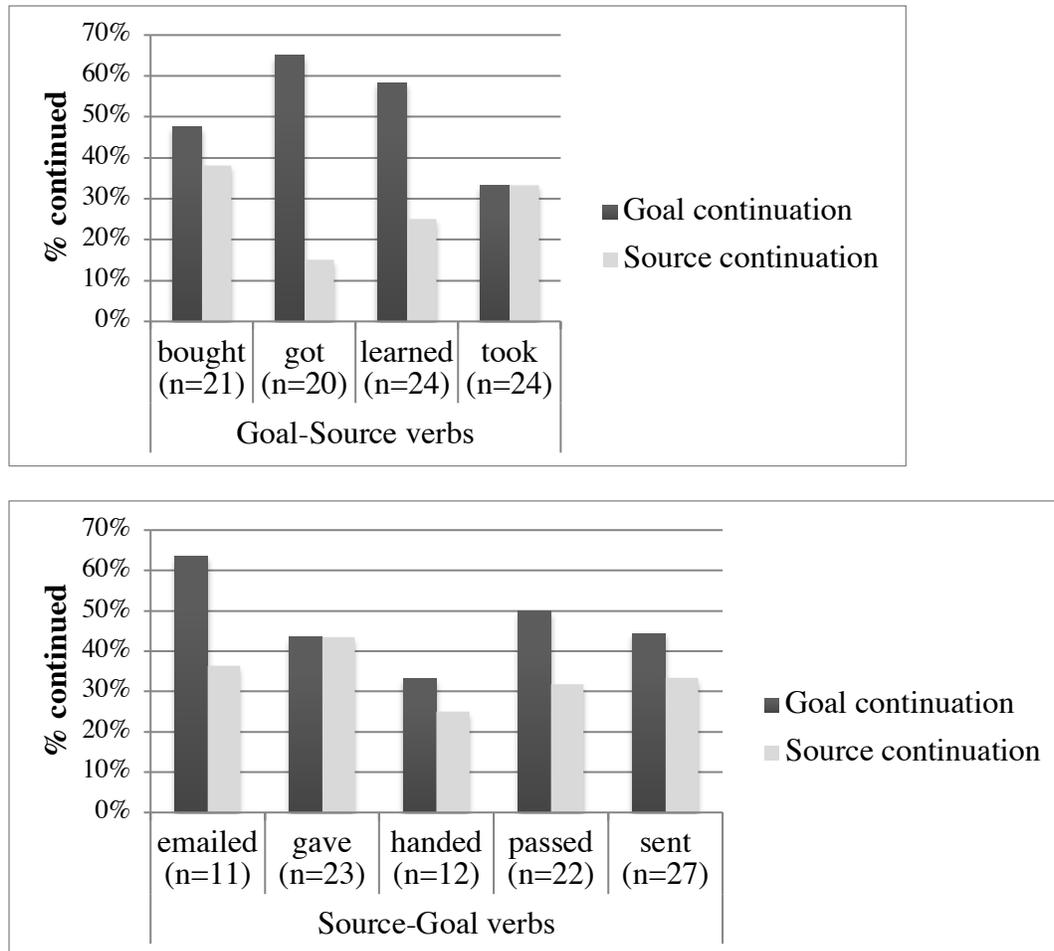


Figure 1. Transfer verbs: Rate of continued mention of goal and source entities. Top panel shows goal-source verbs; bottom panel shows source-goal verbs. The percentage continued reflects the percentage of tokens in which the goal or source is mentioned in the immediately next clause, either directly or indirectly.

Coherence relations

Several theories suggest that the predictability of thematic roles is conditioned on the coherence relations between utterances (Arnold, 2001; Kehler & Rohde, 2013; Kehler, et al., 2008; Rohde, Kehler, & Elman, 2006). In particular, transfer verbs are claimed to focus attention on the goal entity when the next sentence describes the result or following event (i.e., Occasion or Result

continuations), This predicts that goal referents might only be preferentially continued in supportive coherence contexts, or perhaps that the pattern would be stronger in these contexts.

To test these questions, we coded the coherence relation between the critical transfer clause and the following clause, using the system devised by Hannah Rohde and Andy Kehler.² We also added the category “other” to capture cases that did not fit into any of the existing categories. The final coherence ratings were done by a single RA (SW), and compared with coherence ratings for subsets of the data by two other RAs (LR, AMF). Cases of disagreement were solved through discussion between the first author (JA) and SW.

The critical question was whether reference continuation patterns would be stronger in the supportive coherence contexts. Thus, for transfer verbs, is goal continuation more likely in result/occasion vs. other contexts? To test this, we collapsed the coherence ratings into binary categories: result/occasion vs. other for transfer verbs, and explanation vs. other for emotion verbs. To test inter-rater reliability of our coding system, SW double-coded a subset of 31 transfer items and 30 emotion items with two other RAs. We compared agreement for our binary categories, and achieved an average of 91% agreement for the emotion verbs, and 90% agreement for the transfer verbs.

Table 3. Coherence relation coding system

Coherence Relation and Definition	Example
<p data-bbox="203 1587 345 1619"><u>Elaboration</u></p> <p data-bbox="203 1654 802 1717">Elaborates on the same event, e.g. how it is carried out or where/when</p>	<p data-bbox="824 1587 1458 1661">I passed them off to Miss Maria the next day when I took ‘em to her.</p> <p data-bbox="824 1686 834 1707">.</p>

² We are grateful to Hannah Rohde for sharing her coding system with us.

<p><u>Explanation</u></p> <p>Explanation about the previous event or general information about the cause of an event</p>	<p>I bought one from there {for there} really for my daughter because she has problems at night.</p>
<p><u>Occasion</u></p> <p>Temporal relation between two sentences where second sentence describes an event that follows the first sentence, but there is no casual relation (if there is, the explanation coding takes precedence)</p>	<p>I passed my allergies on to them and now I don't have them as bad as they do.</p>
<p><u>Parallel</u></p> <p>Similar event with different referents or similar referents in parallel event</p>	<p>I learned a little bit from my doctor and {I} I read {I- ah} a little bit d coming back. I don't like the bug things.</p>
<p><u>Result</u></p> <p>Causal result of previous event</p>	<p>My husband {ha-} just handed it off to me so i didn't know that was the topic.</p>
<p><u>Violated Expectation</u></p> <p>An unexpected outcome given general real-world knowledge about likely events and their typical consequences/reactions</p>	<p>We sent {um you know} a lot of food and stuff over to Iraq and Afghanistan and they were still against America</p>
<p><u>Background</u></p> <p>Background information that elaborates on some aspect of the event</p>	<p>They handed out these pills to all the postal workers. Yeah they've all got it.</p>
<p><u>Other</u></p> <p>No specific relationship with prior event, or both are related to some higher-level event, or the following clause refers to the entire previous statement.</p>	<p>A: he passed it (= information about the study) on to me. B: Well have you enjoyed it (=the study)?</p>

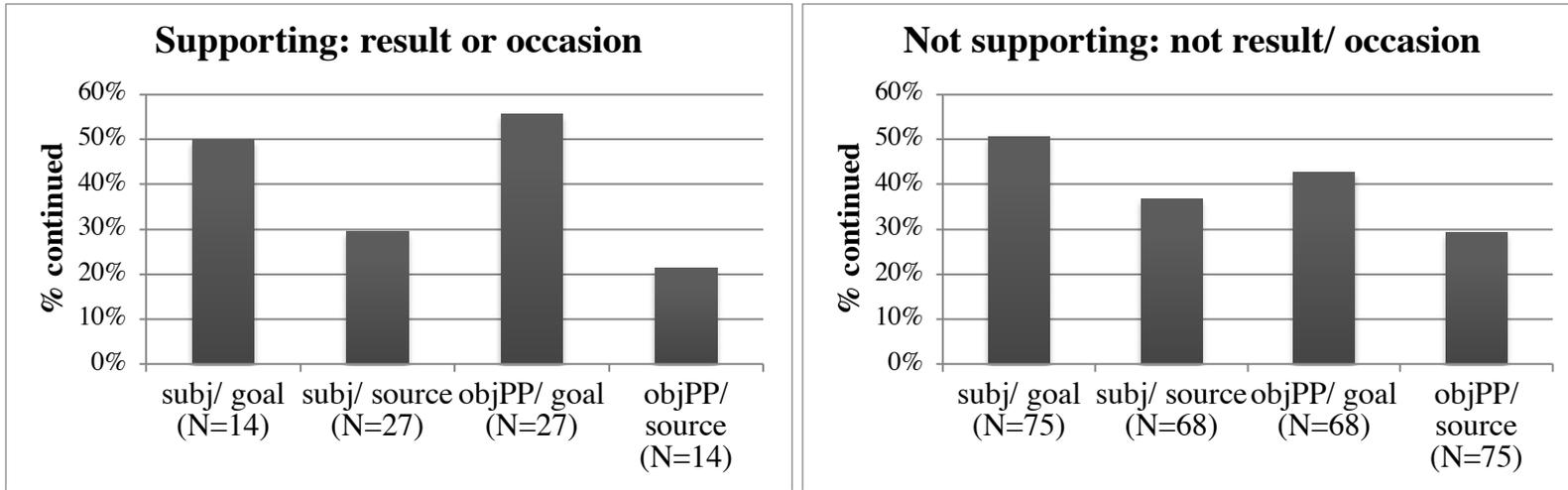


Figure 2. Transfer verbs: The rate of continuation by thematic role and grammatical function, divided by coherence relations predicted to support goal predictability or not.

Figure 2 shows that, contrary to predictions, coherence relation has no effect on the rate of re-mention of goal roles. For transfer verbs (Fig. 1), goals are more likely to be continued, regardless of coherence relation. These patterns are supported by logistic regressions, in which we tested the effect of verbytype, coherence relation, and the interaction between them. As shown in Table 4 there was no interaction for any model.

Table 4a and b. Inferential statistics for the effect of coherence relations and verbytype.

4a. Subject references

Effect	Estimate (SE)	t value	p value
Verbytype	0.64 (0.32)	2	0.082
Causal coherence	-0.18 (0.38)	-0.46	0.645
Verbytype x Causal coherence	0.31 (0.77)	0.4	0.69

4b. Object-of-PP references

Effect	Estimate (SE)	t value	p value
Verbytype	-0.79 (0.32)	-2.46	0.015
Causal coherence	0.07 (0.41)	0.16	0.875
Verbytype x Causal coherence	-0.94 (0.84)	-1.12	0.263

Author note

This work was funded by NSF grant # 1348549 to Jennifer E. Arnold. We are grateful to the following people for their work on this project: Michaela Neely, Grant Huffman, Elise Rosa, Simon Wolf, Leela Rao, and Ana Medina Fetterman.