

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Verb Frequency Explains the Unacceptability of Factive and Manner-of-speaking Islands in English

Permalink

<https://escholarship.org/uc/item/2zb7m1zr>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors

Liu, Yingtong

Ryskin, Rachel

Futrell, Richard

et al.

Publication Date

2019

Peer reviewed

Verb Frequency Explains the Unacceptability of Factive and Manner-of-speaking Islands in English

Yingtong Liu^{1,2} (y_liu@g.harvard.edu), Rachel Ryskin^{2,4} (ryskin@mit.edu), Richard Futrell³ (rfutrell@uci.edu), Edward Gibson² (egibson@mit.edu)

¹ Department of Linguistics, Harvard University, Cambridge, MA 02138 USA

² Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139 USA

³ Department of Language Science, University of California, Irvine, CA 92697 USA

⁴ Department of Speech, Language, and Hearing Science, Boston University, Boston, MA 02215 USA

Abstract

The unacceptability of wh-extraction (e.g., question formation) out of certain syntactic structures, known as ‘island’ effects, has been a central topic in theoretical syntax for many years (Ross, 1967; Chomsky, 1973). A prominent example of islands is that extraction out of a sentential complement introduced by factive and manner-of-speaking verbs (‘What did John *know/whisper* that Mary bought?’) is less acceptable than extraction from a clause introduced by ‘bridge’ verbs (‘What did John *say* that Mary bought?’). We aimed to replicate Ambridge and Goldberg (2008) who argued that extraction from a sentential complement is unacceptable in proportion to its discourse salience. We failed to replicate their results and found that there is no true island effect for such structures: instead there are separate, additive penalties based on two factors: (a) verb-frame frequency (cf. Dabrowska, 2008), and (b) the presence of extraction. These penalties give rise to apparent island effects as a result of the nonlinear relationship between true acceptability and acceptability ratings as measured in Likert scales and forced-choice tasks.

Keywords: Sentence Processing; Frequency Effect; Acceptability of Sentences; Long-distance Dependencies

Introduction

An important feature of human languages is that they contain constructions that license long-distance dependencies--so-called ‘filler-gap’ constructions, such as wh-questions, relative clauses, clefts and topicalization, among others. For example, the declarative form of a simple clause is provided in (1a), along with a wh-question version of this information in (1b), where the patient (object) is extracted. A corresponding relative clause is provided in (1c) and a cleft is in (1d)¹:

- (1) a. Mary bought some apples.
b. wh-question: What_i did Mary buy _i ?
c. relative clause: The apple that_i Mary bought _i
d. cleft: It was the apple that_i Mary bought _i

¹ Following standard notation in the linguistics literature, we will notate the position in the declarative that is associated with fronted element with an empty element ‘ ’. We provide a subscript such as ‘_i’ to the fronted element (the ‘filler’) and the empty position.

Some long-distance extractions are allowed (1), but others are not (2)&(3) (Ross, 1967):

(2) a. * What_i did [S you hear [NP the statement that Jeff baked _i]] ?

b. * Who_i do [S you think [NP the gift from _i] prompted the rumor] ?

(3) (relative clause versions of 2) :

a. * The bread that_i [S you heard [NP the statement that Jeff baked _i]]

b. * The politician who_i [S you think [NP the gift from _i] prompted the rumor].

The unacceptable versions in (2) and (3) have been called ‘islands’ to extraction: unacceptable long-distance filler-gap constructions. The major theoretical interest in island phenomena began with Chomsky (1973), where he argued for a pure structural account, *Subjacency*: noun phrase (NP) and sentence (S) syntactic nodes are bounding nodes for extraction. Extraction across two bounding nodes was proposed to be ungrammatical. Consequently, extractions across the NP and S nodes in (2ab) and (3ab) result in an unacceptable form. Furthermore, Chomsky argued that these constraints must be innate and unlearnable, because (a) the unacceptable extractions occur independent of the meaning of the constructions involved; and (b) a child would not be exposed to the right input across all the different constructions in which they hold (see Schütze et al., 2015, for a summary).

In this paper we focus on extractions out of sentence-complements (S-complements) of factive and manner-of-speaking verbs, as in (4). Researchers have long noted that extraction out of sentence-complements taken by factive verbs – such as ‘know’ (4b), ‘regret’, and ‘notice’, whose S-complements are presupposed (Kiparsky and Kiparsky, 1971) – and manner-of-speaking verbs – such as ‘whisper’ (4c) ‘mutter’, and ‘mumble’, which describe physical characteristics of the speech act (Zwicky, 1971) – are less acceptable than extraction out of ‘bridge’ S-complement taking verbs (4a) (e.g., Erteschik-Shir, 1973; Snyder, 1992; Ambridge & Goldberg, 2008). Note that the definition of a ‘bridge’ verb is not independently defined. A bridge verb is simply one for which extraction from its S-complement is

possible – such as “say” or “think”, which makes the ‘bridge’ baseline of the previous accounts unclear. That is, the notion of ‘bridge’ is not defined in terms of the meaning of the verb, and thus immediately calls into question a potential meaning basis for an observed difference.

(4) a. Bridge verb

What did John **say** that Mary bought?

b. Factive verb

??What did John **know** that Mary bought?

c. Manner-of-speaking verb

??What did John **whisper** that Mary bought?

Previous and Current Theories:

Syntactic Accounts: In order to explain the difference between extraction across bridge verbs (4a) on the one hand and extraction across factive and manner verbs (4b/c) on the other, a syntactic account requires different syntactic structures for bridge verbs compared to the other two kinds of verbs. For instance, it has been claimed that bridge verbs take embedded clauses as arguments, while embedded clauses of manner-of-speaking verbs and factive verbs contain extra covert structures at an abstract level (‘deep structure’ of the Chomskyan framework), such as an invisible complex NP (Kiparsky & Kiparsky, 1971; Snyder, 1992; Stowell, 1981; Stoica, 2016). In this way, the unacceptability of extraction across factive and manner-of-speaking verbs could be captured by syntactic constraints of extraction such as *Subjacency*. However, there are no independent reasons to propose these covert complex structures.

Discourse Accounts: The fundamental idea of discourse accounts is that grammatical constructions specify certain parts of a sentence as ‘foreground’ or ‘background’, and the gap in a filler-gap construction can’t fall within a backgrounded domain. In this spirit, Ambridge & Goldberg (2008) (henceforth A&G) proposed an account they call Backgrounded Constituents are Islands (BCI), arguing that extraction from an S-complement is unacceptable in proportion to its ‘backgroundedness’. The more backgrounded the embedded clauses, the less acceptable the extraction.

Frequency Accounts: Frequency accounts link extraction difficulties to low exposure: less frequent or unpredictable extractions are more difficult to process (cf. Hale, 2001; Levy, 2008). Dabrowska (2008) proposed that speakers store prototypical templates corresponding to frequent combinations they have encountered in their experience such as ‘Wh-word *do you think/say* S-complement?’. Filler-gap constructions that are more similar to the prototypical constructions are more acceptable.

We will propose a different generalization of Dabrowska’s account, following the results of Exp 1 (presented below):

The verb-frame frequency hypothesis: The acceptability of an utterance is best captured by 2 independent, separate effects: (i) the frequency or the type of the construction (e.g., wh-questions vs. declaratives) and (ii) the frequency of the verb head-structure.

Extractions are rated less acceptable than declaratives, because extractions are less common compared to declaratives in communication, or they require more cognitive resources. As for acceptability variance within declaratives or extractions out of S-complements, the major determining factor is the frequency of the matrix verbs taking S-complements (P (matrix verb, S-complement)). This account does not predict an interaction (‘island’) effect between the acceptability of declaratives and extractions. (The interaction obtained in previous works may be a result of applying linear models to non-linear acceptability.)

Following this new *verb-frame frequency hypothesis*, manner-of-speaking wh-questions such as (4c) are less natural, because the manner verbs rarely take S-complements. Factive verbs that take S-complements with a similar frequency to bridge verbs should form equally good wh-questions. A major outlier to our account, the verb ‘*know*’, may be explained by pragmatic factors.

Predictions of The Three Theories on Factive and Manner-of-speaking Islands:

Prediction of the Syntactic Accounts: All factive and manner-of-speaking wh-questions should be less acceptable than all the bridge ones due to categorically distinct covert structures (e.g., Kiparsky & Kiparsky, 1971; Snyder, 1992), as in Fig. 1a.

Prediction of the BCI Account (A&G 2008): There should be a correlation between the acceptability of wh-questions and the backgroundedness of the S-complements taken by the verbs, as shown in Fig. 1b. Factive verbs take presuppositions, the most backgrounded constituents, and therefore should form the most unnatural wh-questions. Manner-of-speaking verbs should form less strong islands, while bridge verbs form fully acceptable wh-questions.

Prediction of Verb-frame Frequency Hypothesis: The acceptability of extraction out of SC verbs should depend primarily on the frequency of those verbs taking S-complements, and the effect of frequency should be similar on both wh-questions and declaratives (no ‘island’ effect), as plotted in Fig. 1c.

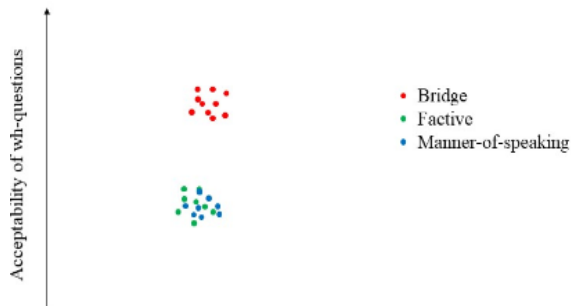


Figure 1a: The prediction of the syntactic accounts

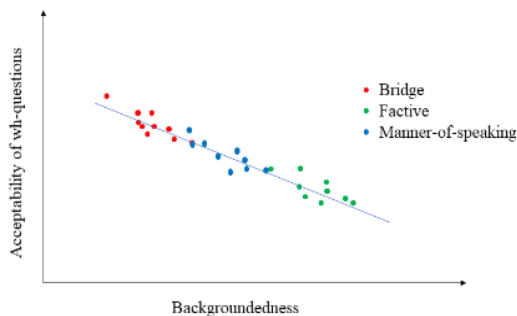


Figure 1b: The prediction of the BCI account.

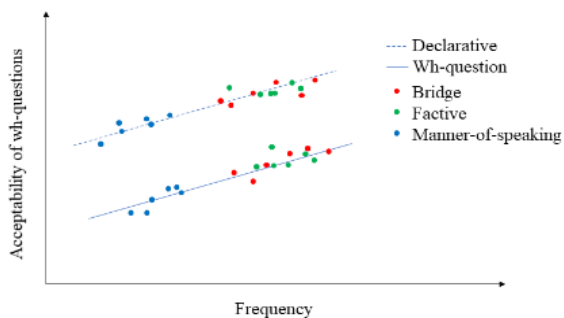


Figure 1c. The prediction of our frequency account.

Experiment 1: Replication of Ambridge and Goldberg (2008)

Experiment 1 is an attempted replication and extension of A&G (2008) using an expanded set of 24 verbs in the 3 categories (A&G tested 12 verbs). There were 2 sub-experiments. Experiment 1a: acceptability judgements of wh-questions formed by the 3 groups of verbs and their corresponding declarative controls. Experiment 1b: negation test to measure the backgroundedness of S-complements of those verbs where extraction appeared. The BCI account predicts a negative correlation between the backgroundedness of the extraction domain and the acceptability of the wh-questions (A&G, 2008).

This experiment also tests the syntactic theories via collecting acceptability ratings of wh-questions formed by the 3 groups of verbs.

Methods

Participants: 180 subjects participated in this experiment via Amazon Mechanical Turk in exchange for \$2 each: 120 participants answered acceptability questions for wh-questions and declarative clauses. Another 60 subjects completed the negation task.

In all the experiments reported here, data from participants who did not self-report themselves as native speakers of American English or didn't answer all the comprehension questions with at least 85% accuracy were excluded. Responses from 116 participants in the acceptability task and 49 participants in the negation task were analyzed.

Design and Materials: The acceptability and negation tasks were constructed for 24 sentence complement (SC) verbs of the 3 categories, as listed in (6)².

- (6) a. Bridge verbs: **say, decide, think, believe**, feel, hope, claim, report, declare
 b. Factive verbs: **know, realize, remember, notice**, discover, forget
 c. Manner-of-speaking verbs: **whisper, stammer, mumble, mutter**, shout, yell, scream, murmur, whine

In the acceptability task, wh-questions and their corresponding declarative sentences were designed as (7a) and (7b) respectively. 96 pairs of wh-questions and declaratives were constructed, and each of the 24 tested verbs in (6) formed 4 pairs. In each pair of wh-question and declarative control, NP1 and NP2 are common names, V1 comes from (6), and V2 is the past tense form of one of the frequently used 25 verbs (*like, eat, buy, build, cook, destroy, dislike, drink, draw, fix, find, know, learn, lose, make, mention, need, see, sell, steal, take, teach, throw, want, write*). To reduce the possibility of semantic plausibility confounds, we used 'something' instead of a specific NP as the embedded object.

- (7) a. What did [NP1] [V1] [[that] [NP2] [V2]]?
 e.g., What did Susan know that Anthony liked?
 b. [NP1] [V1] [that] [[NP2] [V2+something]]
 e.g., Susan knew that Anthony liked something

The 96 pairs were split across 2 lists: each list contained 2 declaratives and 2 different wh-questions per verb. Each participant saw 96 sentences (from 1 list) in a random order. They were asked to rate how natural each sentence was with a rating scale from 1 (extremely unnatural) to 5 (extremely natural). Each sentence was followed by a comprehension question about the content of the preceding sentence to check if participants were paying attention to the task.

In the negation-test task (from A&G, 2008), each trial included a negated complex sentence (8a) and a negated

² Verbs in bold are those tested in A&G (2008). The labeling of a verb as 'bridge' was obtained from previous literature, such as Erteschik-Shir (1973), Snyder (1992), A&G (2008).

simple sentence (8b) which is the negated version of the S-complement in (8a).

- (8) a. [NP1] didn't [V1] [that] [NP2] [V2+Appropriate NP]
e.g., Susan didn't know that Anthony liked the cake.
- b. [NP2] didn't [V2+Appropriate NP]
e.g., Anthony didn't like the cake.

Participants were asked to rate how true they thought the second sentence was, given the first sentence, with a scale from 1 (false) to 5 (true). A&G proposed that these negation scores reflect how “backgrounded” the information in the S-complement is.

Results and Discussion:

A&G (2008) calculated the difference scores between the ratings of wh-questions and declarative clauses as the measurement for acceptability of those wh-questions, and they found a strong correlation between these difference scores and negation scores ($r=-0.83$, $p<0.001$; see Fig.2a). We applied the same analysis to our data. The obtained correlation in our data was in the same direction as in A&G (2008) but the effect was smaller both in the 12 verbs they tested ($r=-0.39$, $p=0.2$) and in the full set of 24 verbs ($r=-0.31$, $p=0.13$; see Fig.2b). Further, we found overlap between acceptabilities for factive and bridge verbs, contradicting the syntactic accounts, which predict non-overlapping acceptability between factive and bridge wh-questions given their distinct covert deep structures.

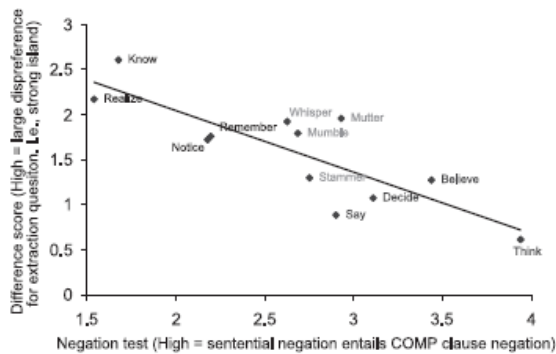


Figure 2a: A&G (2008) - correlation between mean difference scores and mean negation test scores by verb (12 verbs)

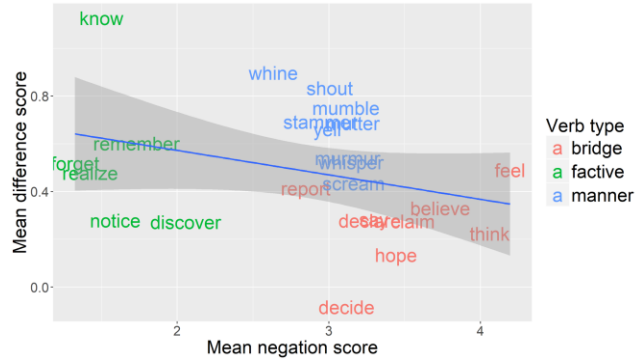


Figure 2b: Our study - correlation between mean difference scores and mean negation test scores by verb (24 verbs).

In a post-hoc analysis, we collected the frequency of the 24 verbs followed by the complementizer ‘that’ from the Google books corpus as a proxy for relative verb frame frequency. Acceptability ratings for wh-question forms were significantly correlated with verb frame frequency ($\rho=0.72$, $p<0.001$), as plotted in Fig.3, as were the corresponding declaratives ($\rho=0.76$, $p<0.001$). Furthermore, 74.6% of ratings were between 4/5 and 5/5 for both the wh-questions and declaratives of verbs, suggesting that participants were not using the full range of the scale.

Thus, we propose *the verb-frame frequency hypothesis*: the acceptability of an utterance is best captured by 2 independent, separate effects: (i) the frequency of the type of construction (e.g., wh-questions vs. declaratives) and (ii) the frequency of the verb head-structure- the frequency of the matrix verbs taking S-complements P(matrix verb, S-complement). This hypothesis suggests the impact of verb frame frequency on filler-gap constructions should be similar to that on declaratives.

An outlier to this account is the verb ‘know’. We hypothesize that the idiosyncratic behavior of ‘know’ was due to pragmatic factors in the wh-question: a question is a request for knowledge but a question with ‘know’ implies that the speaker already has the knowledge. We hypothesize that ‘know’ might not be an outlier in other extraction constructions whose meaning does include implicit knowledge of the interlocutor, such as clefts, which is tested in Exp3.

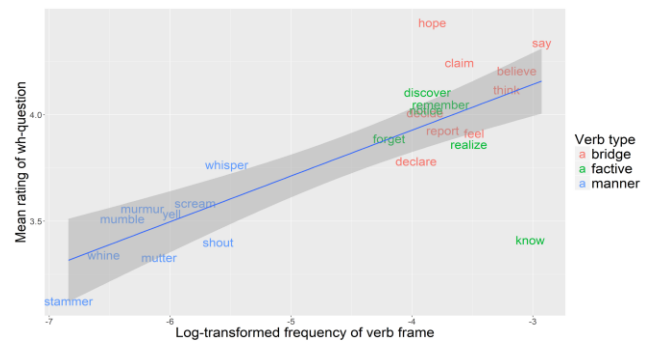


Figure 3: Correlation between mean ratings of wh-questions and log-transformed frequencies by verb ($\rho=0.72$, $p<0.001$).

Experiment 2: Wh-questions with 48 Verbs

The goal of this experiment was to test the frequency account with more matrix verbs beyond the 3 categories (bridge, factive, manner-of-speaking). The *verb-frame frequency hypothesis* predicts that the verbs that frequently take S-complements would be more acceptable as wh-questions and declaratives, regardless of the verb category. The syntactic accounts make no predictions for verbs outside the 3 categories.

Given that the 5-point scale does not seem to be appropriate for measuring the acceptability of these sentences, we performed a forced-choice binary acceptability judgment task in this experiment and applied mixed-effects logistic regression to the data.

Methods:

Participants: 120 people participated via MTurk, and each was paid \$2. Responses from 110 participants were included in the analysis.

Design and Materials: The design was similar to Experiment 1a, with 48 verbs. The verbs included 8 for each of the 3 categories and another 24 outside the 3 categories, as listed in (9). The 24 ‘other’ type verbs were not clearly categorized in the previous literature.

(9) Matrix verbs:

Bridge (8): feel, say, believe, hope, think, report, declare, claim,

Factive (8): know, remember, realize, notice, discover, forget, learn, hate

Manner (8): whisper, mumble, murmur, mutter, whine, shout, yell, scream

Other (24): hear, recall, blab, conjecture, conceal, proclaim, hint, remark, infer, confirm, deny, guess, confide, maintain, testify, reveal, suspect, verify, prove, insist, guarantee, presume, hypothesize, complain

Wh-questions and declaratives were constructed for the 48 matrix verbs with 6 items for each verb (288 items in total). The design of the items is the same as Experiment 1a.

As in Expt 1, participants were assigned to 1 of 2 lists made up of 3 declaratives and 3 wh-questions for each of the 48 verbs. Each participant saw 288 sentences in a random order. Participants were asked to rate each sentence using a binary scale (acceptable vs. unacceptable) based on how natural they thought the sentence was. Each sentence was also followed by a comprehension question.

Results and Discussion:

Acceptability judgments were analyzed with a mixed-effects logistic regression using the *lme4* package in R. *Sentence type* (declarative vs. wh-question), *log-transformed frequency of the verb frame* and their interaction were entered as predictors. The model was fit with the maximum random effect structure which contained random by-*subject*

and by-*verb* intercepts as well as slopes for *sentence type*frequency* by-subjects and slopes for *sentence type* by-*verb*.

The results were in line with the *verb-frame frequency hypothesis*. Wh-questions and declaratives formed by verbs that frequently take S-complements were significantly more acceptable ($\beta=0.58$, $z=3.98$, $p<0.001$). There was also a significant main effect of sentence type: declaratives were rated more acceptable than wh-questions ($\beta=-3.27$, $z=-2.924$, $p<0.004$). No interaction was found ($p>0.4$), suggesting no island effect was present. The log-odds of an ‘acceptable’ response for a given verb frame frequency are plotted in Fig.4a. Note that an island theory would predict the effect of frequency would have a steeper slope for wh-questions than declaratives, but Fig.4a shows the opposite (non-significantly). A pattern resembling a spurious interaction (‘island’ effect) shows up when log-odds are converted into probabilities of acceptance, as shown in Fig.4b.

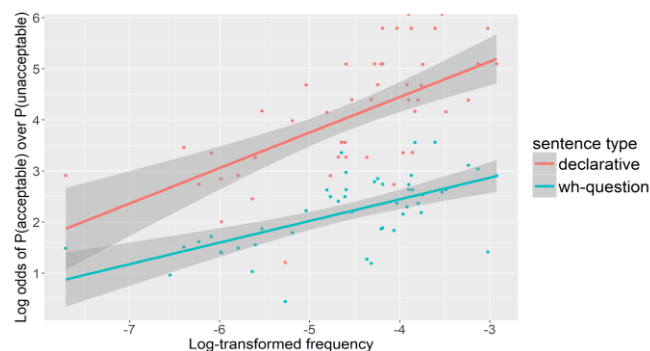


Figure 4a: Log-odds of ‘acceptable’ response for wh-questions and declarative clauses against log-transformed frequencies by verb (48 verbs).

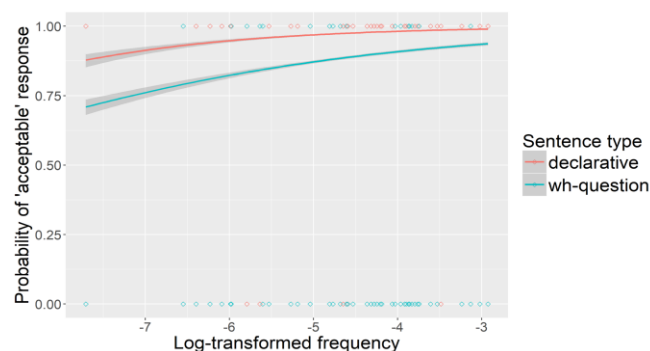


Figure 4b: Probability of ‘acceptable’ response for wh-questions and declaratives against log-transformed frequencies by verb (48 verbs)

Experiment 3: Cleft Structure

Experiment 3 aims to further test the *verb-frame frequency hypothesis* and check if ‘know’ is always idiosyncratic in filler-gap constructions with respect to the frequency account. The syntax-based theories claimed that extractions obey the same set of constraints regardless of construction, which indicates extraction difficulties across different verbs should be the same across constructions (e.g., in wh-

questions and cleft structures). However, an alternative is that the unusual behavior of ‘know’ in Experiment 1 might be related to the ‘information-obtaining’ function of wh-questions. If so, then ‘know’ should not be an outlier in cleft structures, because cleft structures are modifications of an NP and not associated with ‘knowing’. We thus propose that, beyond verb frame frequency, extraction difficulties may differ depending on the meaning and function of the specific construction (Abeillé et al., 2018).

Methods:

Participants: Data from 120 participants were collected via MTurk, and each was paid \$2. Responses from 104 participants were analyzed.

Design and Materials: Cleft structures and their corresponding declarative sentences were designed as in (10a) and (10b) respectively. 96 pairs of clefts and declaratives were constructed, and each of the 24 tested verbs in (6) formed 4 pairs as in Exp 1a. Participants were asked to rate each sentence with a binary rating scale. Each sentence was followed by a comprehension question.

- (10) a. It was the pie that Angela mumbled that Kevin liked
 b. Angela mumbled that Kevin liked the pie.

Results and Discussion:

Acceptability responses were analyzed in the same way as in Exp 2. Sentences with higher frequency verb frames were significantly more acceptable ($\beta=1.2$, $z=2.7$, $p<0.01$) and cleft structures were less likely to be acceptable ($\beta=-14.6$, $z=2.5$, $p<0.011$). The interaction of sentence type and frequency was not significant ($\beta=-0.87$, $z=-0.9$, $p=0.34$), thus providing no evidence of an island effect (Fig. 5). These data are best explained by positing that verb frame frequency and extraction have independent, additive effects in log-odds space, as predicted by the *verb-frame frequency hypothesis*.

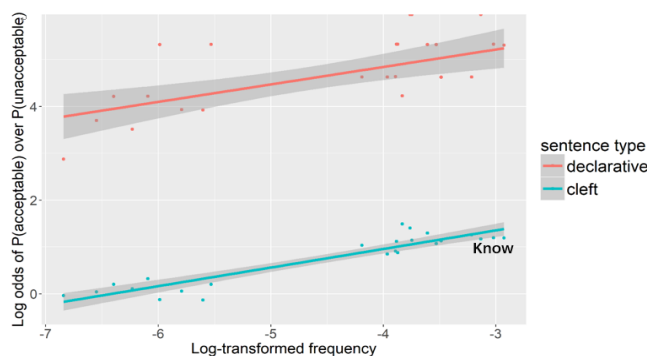


Figure 5: Log-odds of ‘acceptable’ response for clefts and declaratives against log-transformed frequencies (24 verbs)

As predicted by the meaning-based approach to long-distance dependency acceptability, ‘know’ is not an outlier for the frequency account in the cleft structure. The

idiosyncratic behavior of ‘know’ seems to have been due to pragmatic factors in the wh-question: a question is a request for knowledge but a question with ‘know’ implies that the speaker already has the knowledge. If the long-distance dependency structure does not involve the meaning of ‘know’ (as in clefts), then extraction out of S-complements of ‘know’ is acceptable. Such distinct behaviors of ‘know’ in wh-questions and cleft structures suggest extractions vary across constructions, due to their meaning differences.

General Discussion

The results of all three experiments show that the amount of exposure is a key determining factor for the acceptability of filler-gap constructions formed by various SC verbs, including factive and manner-of-speaking verbs. The apparent interaction (‘island’ effect) may be a false positive caused by the use of linear models with ordinal acceptability ratings.

Interestingly, we also found that island constraints are not the same across constructions. Though different extractions may share similar cognitive processes, variation across constructions does exist and can be attributed to different meanings or functions associated with those different types of extractions. Though we didn’t find strong evidence for the discourse-based accounts in the phenomena investigated here, frequency and discourse accounts are not necessarily mutually exclusive in capturing filler-gap constructions (and other phenomena) in general (Abeillé et al., 2018).

Our results suggest that (un)acceptable filler-gap constructions could potentially be learnable via exposure. Although direct negative evidence is missing especially for such complex structures, it is likely that children could use indirect negative evidence to acquire long-distance dependencies. Children may draw statistical inference from the input and regard the absence of a type of extraction in the input as evidence of its unacceptability or ungrammaticality.

Reference

- Abeillé A., B. Hemforth, E. Winckel, E. Gibson. (2018). A construction-conflict explanation of the subject-island constraint, CUNY Conference, UC Davis
- Ambridge, B., & Goldberg, A. (2008). The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics*, 19, 349–381
- Chomsky, N. (1973). Conditions on transformations. In S. Anderson, & P. Kiparsky (Eds.), *A Festschrift for Morris Halle*. New York: Holt, Rinehart, & Winston.
- Dabrowska, E. (2008). Questions with long-distance dependencies: A usage-based perspective. *Cognitive Linguistics*, 19(3), 391–425.
- Erteschik-Shir, N. (1973). *On the nature of island constraints*. Ph.D. dissertation, MIT, Cambridge, MA.
- Hale, J. (2003). *Grammar, Uncertainty and Sentence Processing*. Ph.D. dissertation, John Hopkins University, Baltimore.

- Kiparsky, P. & Kiparsky, C. (1971). Fact. In M. Bierwisch & K. Heidolph (Eds.), *Progress in Linguistics*. The Hague: Mouton
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.
- Ross, J. R. 1967. *Constraints on variables in syntax*. Ph.D. dissertation, MIT, Cambridge, MA.
- Snyder, W. (1992). *Wh-extraction and the lexical representation of verbs*. Unpublished manuscript, MIT, Cambridge, MA.
- Zwicky, A. M. (1971). In a manner of speaking. *Linguistic Inquiry*, 11, 223-233.