Bailey, Laura R., Cook, Eleanor and Kim, Christina S. (2025) *A snapshot of the emerging because-X construction.* **In: English Sociosyntax: Theory, Evidence, Approaches. Topics in English Linguistics, 116 . De Gruyter, Berlin, pp. 227-256. ISBN 978-3-11-142552-8.**

Laura R. Bailey, Eleanor Cook and Christina S. Kim

A snapshot of the emerging *because*-X construction

Abstract: This article gives a description of the emerging construction known as *because*-X (*We have to collect data because science*). Using a Twitter corpus and an acceptability rating study with a large number of participants, we characterise the element X as being sensitive to syntactic category (interjections and bare nouns are favoured, while prepositional phrases and verbs are rarely permitted) and as being syntactically inflexible, unlike the clause or phrasal complement of *because*-CLAUSE or *because*-*of*. Furthermore, we show that there are two subtypes of *because*-X, one with a deictic function indicating a state of affairs, and one with an expressive function conveying an emotional state. Both rely heavily on the addressee's contextual knowledge or salient properties of X, leading to a high degree of flexibility of the interpretation of X and restricted contexts of use in terms of audience (in-groups) and register (informal, especially online communities).

Keywords: *because*; internet language use; register variation; language change

1 Introduction

This paper provides an in-depth discussion of the English construction known as *because*-X (an early observation appears in Bailey 2012). This construction has received attention in the sociolinguistics, corpus and pragmatics literatures (Kanetani, 2015; Rehn, 2015; Bland et al., 2016; Bohmann, 2016; Bergs, 2018; Konvička, 2020). Here, we describe some syntactic characteristics of *because*-X, and present the results of an acceptability survey, with data from a large sample of medium- to high-frequency internet users, as well as a small Twitter (X)[1]

---

1 The social media platform Twitter was renamed 'X' in 2023 after the company was acquired by Elon Musk. We continue to refer to it as Twitter in this chapter, as it was the name by which it was known at the time of the research and as it is generally still known by many people. In part, this is to avoid confusion with the element following *because* in the *because*-X construction.

Laura R. Bailey, Eleanor Cook and Christina S. Kim, University of Kent, e-mail: l.r.bailey@kent.ac.uk

corpus study. We characterise canonical cases of *because*-X as semantically conveying the same causal meaning as standard forms of *because*, but syntactically fixed, with specific conditions of use that limit the audiences and contexts in which *because*-X can be used.

1.1 The *because*-X construction

We distinguish the following three types of *because* construction:[2]

(1) *because*-CLAUSE, a subordinating conjunction (Huddleston and Pullum, 2002, 1020)
We did it because we felt it our duty.

(2) *because*-of, a 'preposition sequence' (Huddleston and Pullum, 2002, 624)
They moved here because of the baby.

(3) *because*-X
There's probably someone out there who would argue the point because internet.

The first two constructions are part of standard Englishes, whereas the last is considered non-standard, though it is widespread in informal usage. Zimmer et al. (2014) trace the origins of the construction tentatively back to a 1991 episode of *Saturday Night Live* and the 90s/00s television series *Buffy the Vampire Slayer*: *I don't even get how we made that guy because, wow, advanced!* Konvička and Stöcker (2022) cite a Dutch example from 1990 and even a very early instance from Nancy Mitford's 1949 novel *Love in a Cold Climate* (*...because, for one things, knee breeches*). It first gained notice among linguists in 2012–2014, with several blog posts noting its existence (e.g., Bailey 2012; Carey 2013; McCulloch 2012, 2014; Garber 2013), and Zimmer et al. (2014) discussing its win as the American Dialect Society's Word of the Year in 2014. Subsequently, Gretchen McCulloch named her book *Because Internet* (McCulloch, 2019), a testament to its status as a salient example of 'internet language'.

Biber et al. (1999) note that adverbials of the *because* type favour clauses (1), and Bohmann (2016) found that of over 12000 instances of *because* in a Twitter corpus, 86% occurred with a clause, 7.7% with an *of* complement (2), and 6.3% with some other type of element. Since instances of *because*-X must be a subset of the last category, the construction is not very frequent. Part of the goal of

---

2 Examples are those provided by Google from Oxford Languages as the search result for the single word *because*.

this study was to investigate the structural and interpretive characteristics of this emerging construction, for those for whom *because*-X is grammatical.

Sometimes known as *because*-NOUN when it was first described, *because*-X does occur with bare nouns as in (3) above. However, it also occurs selectively with some other categories, as early online commentators noted (Bailey, 2012; Carey, 2013; Garber, 2013; McCulloch, 2012, 2014). Carey (2013) gives the following examples from Twitter, among others, of a verbal element (4-a) and an adjective (4-b) as X.

(4) a. Bye going to study for English because didn't finish this morning because fell asleep
     b. Going to bed way early because exhausted :/

In fact, X can be filled by any number of non-linguistic or non-lexical things. (5) gives examples with emojis, real world gestures/actions, and prosodic tunes (linguistic, but lacking any lexical content).[3] (5-a)–(5-d) are real examples from our corpus (see Section 3), and (5-e) is constructed based on similar examples we have witnessed or used ourselves.

(5) a. Y'all I'm late to the party because 😍 🥰 [link]
     b. Who did this bc 😹 😹 [link]
     c. I think in the #MissRwanda contests, all the contestants are crowned.. including the Judge....cos eeeeeehhhhhhhh [link]
     d. If barca wins la liga, I'm sports betting again cus sheeesh [link]
     e. I'm handing this over to them because {<LHLH* tune>, 💁, <shrug>}

These *because*-X examples resemble 'trailing-off' sentences like (6), where the part following *because* does not seem to interact syntactically with the rest of the sentence.

(6) I doubt it because, well, you know.

Adams (2017) suggests that the markers such as *hey* and *wow* in the early examples and *well* in (6) serve as a pivot to a more informal register where slang is acceptable. He notes that in the early instances of *because*-X, adjectives were the most frequent category because they were the most 'unobtrusive' in a 'disruptive' linguistic usage, and that this shifted over time to settle on the later more

---

3 We use angled brackets < > to indicate a physical gesture or action. '[link]' indicates a link included in the original tweet.

frequent *because*-NOUN once the set phrase *because reasons* became popular (the high acceptability of this specific example is shown in Bland et al. 2018). Konvička and Stöcker (2022) similarly trace this pathway from *because*-ADJECTIVE to *because*-NOUN.

The syntactic 'inertness' of *because*-X sets it apart from the standard *because*-CLAUSE, where *because* introduces a subordinate clause. The clause introduced by standard *because* exhibits 'main clause phenomena': of Hooper and Thompson (1973)'s 22 main clause phenomena, standard *because* can occur in all but one context (subject-auxiliary inversion, which renders the clause type incompatible with the cause relation). In standard *because* clauses, reflexive pronouns are marginally licensed with an antecedent in the main clause (dependent on variety), pronouns may corefer with a pronoun or proper noun in the main clause, and proper nouns are not licensed with an antecedent in the main clause. The opposite judgements hold for *because*-X, as shown in (7)–(10).

(7)   a.   The Prime Minister$_i$ wouldn't cut benefits because it would also harm himself$_i$.
        b.   *The Prime Minister$_i$ wouldn't cut benefits because of himself$_i$.
        c.   *The Prime Minister$_i$ wouldn't cut benefits because himself$_i$.

(8)   a.   Every lecturer$_i$ is fond of the student who succeeds because they$_i$ supported them.
        b.   Every lecturer$_i$ is fond of the student who succeeds because of them$_i$.
        c.   *Every lecturer$_i$ is fond of the student who succeeds because them$_i$.

(9)   a.   Every lecturer$_i$ succeeds because their$_i$ efforts are rewarded.
        b.   Every lecturer$_i$ succeeds because of their$_i$ efforts.
        c.   *Every lecturer$_i$ succeeds because their$_i$ efforts.

(10)   a.   Zendaya greets the Spiderman fans who like the film because *Zendaya/she is in it.
        b.   Zendaya greets the Spiderman fans who like the film because of *Zendaya/her.
        c.   Zendaya greets the Spiderman fans who like the film because Zendaya/*her.

There are other indicators that *because*-X is syntactically independent of the main clause. For instance, Bailey and Seyerle (2019) found that the negative scope ambiguity characteristic of standard *because*-CLAUSE was not present in *because*-X constructions. Regardless of the unlikelihood of the negation being interpreted in the main clause, this is the only interpretation for *because*-X:

(11)   George doesn't starve his cat because ethics.
        =*George starves his cat, and the reason is not ethics.
        =George doesn't starve his cat, and the reason is ethics.

(12)   George doesn't starve his cat because money.
        =*George starves his cat, and the reason is not money.
        =George doesn't starve his cat, and the reason is money.

<div align="right">(Adapted from Linebarger 1987, 333)</div>

There are almost no instances of inflection in attested examples (and see Section 2.1 for judgements). Mirroring what Konvička and Stöcker (2022) found in their German and Dutch data, only four verbs were collected in our corpus (see Section 3) and all were participial *-ing* forms (e.g., *subverting expectations*, *according to u and ur comrades*). Definite determiner phrases are also not readily found in the construction as noted by Blamire (2016).

In this article we discuss the construction in English, but it should be noted that it exists in other languages, including French (using the English word) (13), Finnish (14), and Dutch (15):

(13)   Il est 10h30, c'est l'heure de la pause café (le troisieme meug depuis ce matin, bicoz réveil difficil, bicoz couche tard, bicoz Villette Sonique).

        'It is 10.30am, it's time for a coffee break (the third cuppa of the morning, **because difficult wake-up, because night owl, because Villette Sonique)**.'

<div align="right">(Duma, 2014)[4]</div>

(14)   ...mutta en   voi koska   koulu
        ...but   NEG.1SG can because school

        '...but I can't because school.'

<div align="right">(Mervi de Heer, p.c.)</div>

(15)   nee want   lowlands
        no   because lowlands

        'No because lowlands.'

<div align="right">(Konvička, 2019)</div>

---

4 Many thanks to Mercedes Durham for help with the translation of this example.

Konvička and Stöcker (2022) list several other languages where there is evidence for the construction (e.g., Korean, Danish, Czech, and Hungarian) and suggest that it may be very widespread cross-linguistically.

In examples like (3)–(5) and (13)–(15), *because* in *because*-X appears to make the same general meaning contribution as in *because*-CLAUSE and *because-of*: it serves as a discourse relation that identifies a cause or reason, on the one hand, and an effect on the other (see e.g., Kehler (2000)'s CAUSE-EFFECT coherence relation). In addition to the canonical causal interpretation of *because*, some uses of *because* appear to serve a particular discourse function, as in (16). We note that even (16) can be paraphrased as something like 'I know that *Harry will be late* is true because I just talked to his wife, who told me so', which retains a cause/reason interpretation for the *because* clause.

(16)   Harry will be late because I just talked to his wife.

                                                    (Thompson et al., 2007, 8)

In this respect, *because* in *because*-X is one of a small number of lexical items that express causal relations, including *since*, *therefore*, and *so* — these differ in terms of which element (cause, effect) appears first. However, in some respects, the standard *because-of* form is more like *because*-X, and unlike its *because*-CLAUSE counterpart. As noted in prior work on *because*-X (e.g., Konvička and Stöcker 2022), *because*-X has in common with *because-of* that it relies on the hearer and speaker having shared knowledge or beliefs about how the nominal complement is relevant to the discourse, or a shared understanding of a relevant quality of the referent (Blamire, 2016). In (17-a), for example, we need to either know or infer something about the cat, so we need to either be aware that his cat meows at dawn or be able to infer that his cat needs him early in the morning for food, because it is a salient fact about cats that they behave this way. Compare the examples in (17) with their counterparts with clausal complements in (18).

(17)   *because-of*:
        a.   He got up early because of the cat.
        b.   She got up early because of her departure time.

(18)   *because*-CLAUSE:
        a.   He got up early because the cat started meowing at dawn.
        b.   She got up early because her plane left at 4.30. (= in order to catch her plane at 4.30)

                    (Adapted from Huddleston and Pullum 2002, Section 12.2)

Rehn (2015) suggests a pragmatic restriction on the reduced form of *because*-X and *because-of* relative to *because*-CLAUSE: the common ground shared by the speaker and hearer must already contain enough information for the hearer to be able to recover the missing content (i.e., material that would appear in a clausal complement), and for the speaker to be able to assume that the hearer will be able to recover the intended meaning.

Finally, a distinguishing characteristic of *because*-X is the sarcasm that is integral to many uses of this construction. We see the sarcastic aspect of the meaning overtly expressed in some of our own corpus data (discussed in full in Section 3): (19) is written in 'SpongeBob' formatting (a mixture of upper and lower case used in a meme based on the cartoon character 'Spongebob Squarepants' for sarcastic or mocking comments).[5]

(19)   bc sUbVErtInG eXpeCtaTiOnS

Note that the meaning of *because*-X can be conveyed by uttering a bare X, without the introductory *because*. This is illustrated in (20), a question from an online form, in which the response *Millennial* has the same interpretation as *Because millennial* would have done.

(20)   A: I can only be contacted by text message (SMS).
         B: Explain why you cannot be contacted by telephone call.
         A: Millennial.

The *because*-CLAUSE counterpart, *Because I am a millennial* or *I am a millennial*, differs not only in register but in lacking the sarcastic meaning present in the *because*-X construction. We return to this point in the General Discussion.

1.2 Aims and outline

Our primary goal in this paper is descriptive: what kinds of expressions can appear as X in *because*-X? Because there is little work on this emerging construction, and speakers' judgements vary widely, it is important to first establish an empirical base with respect to the well-formedness of a range of forms of *because*-X, for those speakers who allow *because*-X at all. We address this descriptive question with a Twitter corpus study and an acceptability judgement survey with a large number of respondents. Because participants were recruited

---

5 The origins of the meme can be found at Know Your Meme (https://knowyourmeme.com/memes/mocking-spongebob).

via social media, they are likely to be regular social media users, and therefore likelier than the general population to have been exposed to usage of *because*-X. While we disagreed with a number of early online commenters that the construction involves ellipsis (see also arguments in Blamire 2016; Konvička and Stöcker 2022), we did expect there to be sensitivity to the syntactic category of X, based on discussions in blog posts and popular media and our own intuitions. Specifically, we expected bare nouns and interjections or 'pivot' items to be the most acceptable, as paradigmatic cases of *because*-X. We expected noun phrases including determiners to be less acceptable (Fortin, 2007; Blamire, 2016), and verbs relatively unacceptable (Konvička and Stöcker, 2022). Our findings indicate that respondents do have grammaticality judgements regarding what syntactic category can occur as X, and that levels of internet usage and native speaker status affect those judgements in a way that is consistent with an emerging construction.

Based on the acceptability patterns observed in the survey, we propose a rough characterisation of the structural and interpretive properties of *because*-X. In particular, we argue that X — instead of being derived from either of the two standard *because* forms by linguistic processes like deletion or reduction — is not necessarily linguistic, as it lacks evidence of internal syntactic structure. We further demonstrate that *because*-X's interpretation differs from those of *because*-CLAUSE and *because-of*, and we propose two subtypes of *because*-X: a deictic subtype that points to some current situation, and an expressive subtype that conveys an emotional state. These subtypes have overlapping characteristics, and need not be exclusive of each other.

Moreover, the use of X as though it were a syntactically appropriate complement of *because* is deliberately ungrammatical in certain defined ways, as is characteristic of other forms of 'internetese' (cf. 'doge' memes and various dialects of cats and dogs online in, e.g., the ongoing work of Edith Podhovnik) and as a form of 'language play' (Gawne and Vaughan, 2011). Related to its reduced form, *because*-X relies more heavily on common ground shared by speakers and addressees than canonical forms of *because*, and is more pragmatically constrained with respect to the conversational circumstances under which it can be used.

2 Preliminary surveys to narrow down the field

Two exploratory surveys were used to narrow down the range of possible forms that X can take in *because*-X, in advance of the survey presented in Section 4. An

acceptability judgement survey asked respondents to rate instances of *because*-X, each featuring a different category in X. A sentence completion, or 'fill-in-the-blank', survey asked what (if anything) was missing from the same instances of *because*-X.

2.1 Acceptability judgments

Since the range of expressions that can appear as X in *because*-X appears to be wider than what is possible for canonical *because* sentences, this exploratory survey sought to narrow down the range of possible forms that X can take.

2.1.1 Method

Twenty-two sentences containing the target construction *because*-X were presented to participants in an online survey. Participants were asked to indicate how natural the sentence sounded on a 1–6 scale (1='really weird', 6='totally natural'). The element X came from eight syntactic categories (e.g., adjective, noun, prepositional phrase), comprising 22 distinct sub-categories (e.g., bare noun, proper noun). The full list of sentences is in Appendix A.

There were five optional demographic information questions, which we do not discuss further beyond noting that younger speakers are more accepting of *because*-X, and that 97.2% of respondents indicated that they were either native speakers of English or were highly proficient in English. The items were attested examples where possible (13 sentences), taken mostly from Carey (2013), with unattested categories created (9 sentences) and matched as closely as possible in terms of register and content. The survey was distributed as a link shared on Twitter. 247 respondents participated voluntarily, and received no compensation.

2.1.2 Results

The mean and median ratings for each category of X are given in Table 1.[6]

---

[6] Because it has been argued that Likert scale data are ordinal and should therefore be analysed using nonparametric methods, we report both the median and the mean throughout.

Category	Rating		
	M	*Med*	*(SD)*
Standard *because* (*because*-CLAUSE *because-of*)	4.9	6.0	(1.71)
Interjection	4.8	5.0	(1.5)
Bare noun	4.2	5.0	(1.8)
Adjective	3.5	4.0	(1.7)
Noun phrase	3.3	3.0	(1.7)
Prepositional phrase	2.9	2.0	(1.5)
Verb phrase	2.9	2.0	(1.7)
Clause with overt complementiser (*that*, *if*)	1.6	1.0	(1.3)

Tab. 1: Mean and median ratings by X category for exploratory acceptability survey. Rating scale: 1–6 (1='really weird', 6='totally natural').

Participants were aware of the topic of the survey and thus tended to be accepting of the construction in general: the most typical example of *because*-X, *because school*, received relatively high overall ratings (M=4.5, Med=5.0, SD=1.7), and 79.8% of the participants gave a rating of greater than 3 across the board.

While ratings ranged over the entire scale, judgements did suggest a sensitivity to syntactic category: not just any category can fill X. Consistent with previous work, bare nouns were among the highest rated, but they were not the only possibility.

2.2 Sentence completion

We conducted an exploratory sentence completion survey, in which respondents were asked to add to *because*-X sentences if they felt anything was missing. This allowed us to get a sense — for speakers for whom *because*-X is *not* grammatical — for what they understood *because*-X to be short for (i.e., what respondents assumed their non-reduced, standard *because* counterparts to be). Because those who found a *because*-X item acceptable were able to respond that they would not change anything, this survey also provides another way of measuring acceptability.

2.2.1 Method

One sentence from each of five categories (shown below in (21)–(25)) was selected and presented to 957 participants via a survey distributed on Twitter. They were given the following instructions: *"If you think anything is missing from the*

sentence, please rewrite the sentence. You can provide more than one option if you want to. If you don't think anything is missing, just type an x into the box."

(21)   Interjection
         what is he thinking, because honestly

(22)   Bare noun
         i'm going to bed because self-care

(23)   Prepositional phrase
         excuse my typos because on mobile

(24)   Determiner phrase
         the quality is really bad because my computer

(25)   Bare verb[7]
         i really need to lose some weight because eat

As no additional context was provided for the sentences, respondents' interpretations relied on their own world knowledge and attitudes. Although there was an option for participants to indicate that nothing was missing, the question may have encouraged people to provide an answer if they could (e.g., if they found the sentence acceptable as is, but could also think of a differently-worded version which was also acceptable). Sentences were presented as in (21)–(25), without gaps/underscores to suggest where material might be inserted.

Participants followed a wide range of strategies, including inserting extra material, rewriting the sentence, or indicating that the sentence was ungrammatical. Responses were coded according to whether the participant indicated that nothing was missing (No change), altered the sentence to a standard form by inserting *of* before a noun phrase or forming a clause (Standardise), or indicated that the sentence was not grammatical either by explicitly saying so or by entirely rewriting it (Ungrammatical). Responses that did not fall into one of these categories (8.9% of the data) were excluded from the analysis.

2.2.2 Results

Table 2 shows the proportion of responses of each type for each of the five categories.

---

7 A number of respondents pointed out that (25) exhibits fatphobic sentiments. While we tried to exclude tweets that could cause distress as we selected stimuli from our Twitter corpus, we apologise for having missed this one.

		Proportion responses		
Category	Item	No change	Standardise	Ungrammatical
Interjection	*because honestly*	.8	.1	.03
Bare noun	*because self-care*	.7	.2	.2
Prepositional phrase	*because on mobile*	.3	.5	.1
Determiner phrase	*because my computer*	.3	.7	.02
Bare verb	*because eat*	.08	.7	.09

Tab. 2: Proportion of responses of each type by category of X for sentence completion survey.

The interjection sentence (21) and bare noun sentence (22) were likeliest to be left unchanged. This is in line with the acceptability judgements described in the previous section. The determiner phrase sentence (24) was most often standardised to *because of my computer*, while the bare verb (25) and prepositional phrase (23) sentences were most often standardised to the *because*-CLAUSE form.

3 Corpus study

In addition to testing people's acceptability judgements of sentences with *because*-X, we collected a small corpus of tweets in order to find out how people use the construction in informal written online language. While there are no prior acceptability studies of *because*-X, Twitter corpora have been used by previous researchers (Schnoebelen, 2014; Bohmann, 2016) to look at the distribution of *because*-X.

3.1 Method

We collected tweets containing the words *because* and its variants *cos*, *cus*, and *bc* using the web-scraping tool Octoparse (Octopus Data Inc., 2023), which automates a search, in January and August 2020 The tweets were hand-sorted to remove standard instances of *because*, leaving a corpus of 706 tweets containing a non-standard use of *because* (full corpus). This subset comprised around 5% of the total tweets containing the word or one of its variants. For comparison with Schnoebelen (2014)'s data, we also created a subset of the tweets with a single word following *because* following the same principles he used (end of tweet, punctuation, emoji, end of the line, or RT), leaving us with a sub-corpus of 254 tweets (1-word corpus).

We coded both the full corpus and the 1-word corpus using the categories in Schnoebelen (2014) (Noun, Compressed clause (e.g., *ilysm* 'I love you so much'), Adjective, Interjection, Agreement, Pronoun). In addition, for the larger non-standard *because* corpus, we added some categories (*what/why*, Nothing — for utterance-final *because*, Prepositional phrase, Verb) that were not present in Schnoebelen's data or were present but excluded from analysis. Note that Schnoebelen only counted lexical items with 50 or more occurrences. We included all tweets irrespective of frequency.

Both Bohmann (2016) and Schnoebelen (2014) classified some complements as 'compressed' or 'reduced' clauses. For Bohmann (2016), these were finite clauses lacking a subject, while for Schnoebelen (2014) (who was restricted to single-word clauses) 'compressed' clauses were utterances such as *yolo* ('you only live once'). We followed Schnoebelen's practice here.

3.2 Results

Table 3 shows the distribution of instances of *because*-X by category of X.

	Full corpus Percent tweets (*n*)	1-word corpus Percent tweets (*n*)
what/why	28.75 (203)	3.94 (10)
Interjection	19.69 (139)	35.83 (91)
Noun	19.26 (136)	21.26 (54)
Nothing	12.18 (86)	0 (0)
Adjective	7.08 (50)	17.72 (45)
Agreement (*yes/no*)	5.67 (40)	12.99 (33)
Compressed clause	2.12 (15)	3.94 (10)
Pronoun	1.56 (11)	1.57 (4)
Emoji	1.27 (9)	2.36 (6)
Adverb *-ly*	0.57 (4)	1.57 (4)
Verb	0.57 (4)	0 (0)
Prepositional phrase	0.28 (2)	0 (0)

Tab. 3: Categories following *because* in our full corpus and the 1-word sub-corpus.

Wh- questions, which Schnoebelen (2014) did not include in his corpus, were frequent in our data. Clearly, an information-seeking question is not suitable as a cause or implied reason, as expected for standard uses of *because*. This points to the use of *because* as a discourse continuation marker or 'paratactic' *because*, which Burridge (2014) analyses as undergoing grammaticalisation. This

discourse marker 'doesn't assert a reason but rather provides the grounds for why a speaker knows or believes what is expressed in a prior clause' (2014, 533). In the classification system used here, some *wh-* expressions (e.g. *what the fuck*) were classified as interjections (in our full corpus, around a quarter of tweets classified as interjections were *wtf* or related expressions), and full clauses (e.g., *because what the fuck is this?*) were classified as *wh-* questions. *Why not* was especially common in this category, and although these were classified as a *wh-* questions because they can serve as such in isolation, we note that these were usually not information-seeking questions (cf. *posting it on here too cus why not*).

Our corpus differs notably from Schnoebelen (2014)'s in having many more instances of interjections as X (Table 4): in the 1-word sub-corpus interjection is the most frequent category. The large number of interjections compared to Schnoebelen's data may be due to our use of the whole sub-corpus as opposed to only those exceeding a frequency threshold, as there was a wide range of highly infrequent items (e.g., *chale* and *fam*, both address terms meaning something like 'mate' or 'buddy', occur once each).

	Schnoebelen (2014) Percent tweets	Our 1-word corpus Percent tweets (*n*)
Noun (*people, spoilers*)	32.02	21.26 (54)
Compressed clause (*ilysm*)	21.78	3.94 (10)
Adjective (*ugly, tired*)	16.04	17.72 (45)
Interjection (*sweg, omg*)	14.71	35.83 (91)
Agreement (*yeah, no*)	12.97	12.99 (33)
Pronoun (*you, me*)	2.45	1.57 (4)

Tab. 4: Comparison of Schnoebelen (2014)'s data with our 1-word sub-corpus.

Verbs were very rare in Schnoebelen's corpus; we had none in the 1-word sub-corpus and only four in the larger corpus, all with *-ing*, which again mirrors Schnoebelen's findings. These participial (*-ing*) forms are nominalisations, with a distribution like that of nouns. As such, they are perfectly grammatical in a *because-of* sentence (*because of jogging*), where only nominal complements are possible. This almost total lack of verbs once again matches the finding of the acceptability rating survey, where they received very low ratings, and the sentence completion survey, where most participants transformed them into a standard *because*-CLAUSE.

3.3 Interim summary

This small scale corpus study is consistent with the results of the exploratory surveys described in Section 2. Interjections and bare nouns are notably the most acceptable in *because*-X, and the most frequent. (We set aside *what/why* for the moment, since *wh*- expressions were sometimes classified as interjections, and sometimes as *what/why*, based on whether the same form could function as an information-seeking *wh*- question in isolation. Most of these *wh*- questions are not information-seeking in context, as noted above, and we return to this in Section 5.2). Verbs appear to be particularly unacceptable and infrequent in *because*-X. While we selected the items used in the initial surveys based on their syntactic categories, we cannot interpret the results as representing restrictions associated with those categories: because the materials included on average fewer than three tokens per syntactic category (the sentence completion survey featured only one item per type, in fact), we cannot generalise beyond those tokens. However, it is suggestive that the distribution of categories in the corpus study is remarkably similar to the pattern of acceptability we see in the exploratory surveys.

4 Acceptability survey

Focusing on the five categories from the sentence completion survey (interjection, bare noun, determiner phrase, prepositional phrase, bare verb), we collected acceptability judgements on a larger set of stimulus items, from a large population representing a range of speaker types and levels of internet usage.

Our aims were to determine whether the patterns of acceptability we observed in the exploratory surveys and the corpus study are replicated with this larger set of materials, and ask to what extent speaker background (native speaker status) and level of internet usage contribute to explaining individual acceptability judgements for *because*-X.

4.1 Method

4.1.1 Materials

Six *because*-X sentences were created for each of the five categories, resulting in 30 items in total. The categories were as follows: interjections (*because ew*), bare nouns (*because misogyny*), determiner phrases (*because the lyrics*), prepositional phrases (*because at this point*) and bare verbs (*because shout*). The determiner

phrases contained a range of determiner types (*the, this*), and the prepositional phrases each used a different preposition. The verbs were all unergative, following arguments in the literature that these are the most acceptable verb type in fragment utterances (Fortin, 2007). Having already seen an indication that transitive verbs are highly unacceptable in Survey 1, we hoped in this way to identify whether verbs are ever rated as acceptable.

The stimuli were all based on existing examples found on Twitter. Where possible, we used a natural example, maintaining the structure but replacing some words with synonyms to maintain anonymity (Durham, 2016; Tatman, 2018). Where we were unable to find an example of a particular category, we minimally edited a naturally-occurring example of one of the other *because* types to create a *because*-X sentence. The full list of sentences is in Appendix B.

The web-based survey was distributed via the first author's social media networks (Twitter, Facebook, and LinkedIn).

4.1.2 Participants

957 respondents completed the survey. They provided informed consent, and did not receive compensation for participating.

After the main acceptability portion of the survey, participants were asked to provide basic demographic information, including what variety of English they spoke (including whether they were a non-native speaker of English) and their level of internet usage. Responses to the English variety question were categorised as one of: North American variety, UK variety, other English variety (e.g., Australian, South African, Singaporean), or non-native speaker. Those who provided their level as C2 or its keyword descriptors ('near native') were classified as North American, UK, or Other English variety. Those who described themselves as 'fluent' were not classified with native speakers.

The internet usage question was included since *because*-X is widely perceived to be 'internetese'. Respondents who mentioned one or two very common platforms such as Whatsapp or specified spending less than one hour a day online were classified as having 'low' usage. Those who mentioned several platforms or mentioned being online 2–3 hours per day were classified as 'moderate' users. We classified as 'high' usage anyone who self-declared as such, who said they spent many hours a day online, or used some kind of meme in their answer. The distributions of English variety and level of internet usage in our sample are shown in Table 5.

	Category	Percentage of sample
English variety	**North American English**	**51.5**
	UK English	**26.6**
	Other English variety	**13.3**
	Non-native speaker	**8.6**
Internet usage	**high**	**44.7**
	moderate	**35.9**
	low	**19.3**

Tab. 5: Distribution of English variety and level of internet usage in our sample (N=957).

4.1.3 Procedure

Respondents rated each of the 30 items on a 1–6 acceptability scale (1='very unnatural'; 6='very natural'). Items were presented in a separate random order for each respondent, and were presented on a single screen. Completing the survey took participants less than 10 minutes each.

4.2 Results

The mean and median acceptability ratings by category are shown in Table 6.

Category	Rating		
	M	**Med**	**(SD)**
Interjection	**4.9**	**5.3**	**(1.1)**
Bare noun	**4.1**	**4.3**	**(1.2)**
Determiner phrase	**3.3**	**3.3**	**(1.2)**
Prepositional phrase	**2.5**	**2.3**	**(0.9)**
Bare verb	**1.7**	**1.3**	**(0.8)**

Tab. 6: Mean and median ratings by X category. Rating scale: 1–6 (1='very unnatural'; 6='very natural').

Because of the large number of participants, our survey is likely overpowered. It is therefore not surprising that the ratings for each of the categories are highly significantly different from the ratings for each of the other categories (for pairwise paired t-tests, all $t>20$, df=956, all adjusted p<.001).

4.2.1 Effects of internet usage, native speaker status

Looking at each category of X separately, we ask whether internet usage level or native speaker status are predictive of acceptability for that category, and whether those factors interact with each other. For each category, we fitted a mixed-effects model predicting acceptability rating, with internet usage, native speaker status, and their interaction as fixed effects, using the maximal random effects structure (Barr et al., 2013). Analyses were performed using R Statistical Software (v4.3.0; R Core Team, 2021) and the `lme4` package (Bates et al., 2015).[8] Coefficient estimates for the five regression models are in Table 7.

Category	Predictor	Estimate	*SE*	*p*
Interjection	(Intercept)	4.88	0.26	<.001
	Internet usage	0.36	0.092	< .001
	Native speaker	-0.066	0.68	.34
	Int usage:Native	0.12	0.084	.4
Bare noun	(Intercept)	3.97	0.33	<.001
	Internet usage	0.47	0.11	< .001
	Native speaker	0.048	0.069	.49
	Int usage:Native	0.063	0.094	.51
Determiner phrase	(Intercept)	3.37	0.18	<.001
	Internet usage	0.36	0.093	<.001
	Native speaker	-0.14	0.085	.10
	Int usage:Native	-0.13	0.092	.16
Prepositional phrase	(Intercept)	2.53	0.35	<.001
	Internet usage	0.029	0.082	.73
	Native speaker	-0.073	0.068	.30
	Int usage:Native	0.022	0.079	.78
Bare verb	(Intercept)	1.77	0.076	<.001
	Internet usage	0.027	0.083	.75
	Native speaker	-0.11	0.056	.049
	Int usage:Native	0.029	0.079	.72

Tab. 7: Coefficient estimates for regression models predicting acceptability, by X category.

---

8 We also analysed the data as ordinal data using cumulative link mixed models, using the `ordinal` package (Christensen, 2023). These models did not differ from the linear regression models presented in Table 7 in terms of the direction or significance of fixed effects terms, with the exception of the determiner phrase model, where the Native speaker predictor reached significance (β=-0.25, SE=0.13, p=.049).

For interjections, internet usage was a strong predictor of acceptability, with higher usage participants giving higher acceptability ratings. There was no effect of native speaker status, and the internet usage effect did not differ by native speaker status. Bare nouns and determiner phrases showed the same pattern, with a positive effect of internet usage, no effect of native speaker status, and no interaction between the two.

There was a reliable native speaker effect — with native speakers having sharper ungrammaticality judgements than non-native speakers — in the bare verb model. Note that verbs were the lowest rated overall, and very infrequent in our corpus. The model also showed no effect of internet usage, and no usage-by-nativeness interaction.

Finally, the prepositional phrase model showed no reliable effects of internet usage, native speaker status, or their interaction.

4.2.2 Within-speaker variability

We looked separately at respondents for whom *because*-X was generally acceptable (*grammatical* group: those whose mean rating for interjections and bare nouns was greater than 5 out of 6), and respondents for whom *because*-X was generally unacceptable across the board (*ungrammatical* group: those whose mean rating for interjections and bare nouns was less than 3 out of 6). The mean and median ratings and mean by-speaker variances by X category and grammaticality group are in Table 8.

Category	Ungrammatical group (n=63)			Grammatical group (n=215)		
	M	*Med*	*(M₂)*	*M*	*Med*	*(M₂)*
Interjection	2.0	2.0	(1.2)	5.7	6.0	(0.52)
Bare noun	1.6	1.0	(0.61)	5.5	6.0	(0.89)
Determiner phrase	1.9	1.5	(0.87)	3.9	4.0	(2.0)
Prepositional phrase	1.9	1.0	(1.3)	2.7	2.5	(2.9)
Bare verb	1.4	1.0	(0.27)	2.1	1.5	(1.0)

Tab. 8: Mean and median ratings and mean by-subject variance by grammaticality group and X category.

For each group, we examined the within-speaker variance in acceptability for the highest-rated (interjection, bare noun) and lowest-rated (prepositional phrase, bare verb) categories. For speakers in the ungrammatical group (n=63),

mean by-speaker variance in acceptability was comparably low for the highest-rated (M_{σ^2}=0.94) and lowest-rated (M_{σ^2}=0.87) categories. This is what might be expected if respondents were making categorical judgements. Even within this group, however, respondents showed some gradience in acceptability judgements.

For speakers in the grammatical group (n=215), mean by-speaker variance in ratings was low for the highest-rated categories (M_{σ^2}=0.68) but substantially higher for the lowest-rated categories (M_{σ^2}=2.2). This difference from the un-grammatical group is suggestive of the fluidity of *because*-X at the moment, even for fluent *because*-X users.

4.3 Discussion

Our survey shows a clear hierarchy of linguistic categories in terms of how acceptable they are in *because*-X. For the three highest-rated categories — interjections, bare nouns, and determiner phrases — acceptability was related to a respondent's level of internet usage: the more 'online' someone was, the more acceptable they found *because*-X with these categories. For the lowest-rated prepositional phrase and bare verb categories, there was no such effect: these instances of *because*-X were rated as highly unacceptable, no matter how much or little a respondent spent time online.

In addition, *because*-V, and *because*-DP, marginally, show native speaker effects, with native speakers rating these as more unacceptable, independently of their level of internet use. We speculate that, as *because*-X gains usage outside of highly-online groups, the larger population of native speakers may be developing grammaticality judgements. In other words, even someone who would not often use or come across *because*-X sentences may have greater certainty that *because*-VERB is 'ungrammatical' as *because*-NOUN becomes more widespread, and sounds more 'grammatical'.

As noted in Section 4.1, respondents were recruited from the first author's social media following, and from retweets by her followers. This introduces an obvious skew in the data presented here: as a number of respondents noted, many respondents had degrees in linguistics or related subjects, and were aware of issues like prescriptivism and language change. Other respondents, on the other hand, made specific mention of grammar rules, or suggested that their responses reflected their adherence to such rules (due to education, what they do for a living, age). It is not the case, however, that those who were more permissive with *because*-X were indiscriminately permissive: just 0.84% of respondents consistently rated the lowest-ranked categories above 4.

5 General discussion

Based on the findings we report in Sections 3–4, we provide a sketch of *because*-X, first by describing what sets it apart from standard *because*, then describing the characteristic properties of *because*-X and its conditions of use.

5.1 What *because*-X is not

Two features of *because*-X make it different from *because*-CLAUSE and *because-of*. First, X in *because*-X appears to be syntactically fixed.

Most obviously, single words such as interjections and bare nouns are more acceptable than multi-word phrases (though we revisit this point in Section 5.2), and the presence of agreement or selection of a complement appear to degrade judgements. There is an almost total lack of prepositional phrases and inflected verb phrases, with the few verbs attested generally being participial forms with no complement. The ultimate expression of the single-item constraint is the acceptability of emoji and gestures, which, while used communicatively, contain no lexical content. These are readily found, and such expressions were offered by some participants as alternatives in the sentence completion survey discussed in Section 2.2.

As discussed in Section 4, determiner phrases are lower rated compared to bare nouns, contrasting with standard *because-of*. Bare nouns are only acceptable in *because-of* contexts if they are also possible in other syntactic contexts, so while *because television* and *because internet* are both acceptable, **because of internet* is not well-formed, compared to *because of television*:

(26)   a.   Television has made him afraid to leave the house.
         b.   *Internet has made him afraid to leave the house.

(27)   a.   He won't leave the house because of television.
         b.   *He won't leave the house because of internet.

Anecdotal data also suggests that singular nouns are preferred to plural nouns: *because exam* is better than *because exams*.

As well as lacking internal structure, the X in *because*-X seems to be independent of the main clause that it is apparently subordinate to. In this it is somewhat similar to standard *because* clauses which exhibit root phenomena, and other (standard) uses of *because* include the paratactic or metalinguistic functions as noted in Section 1. However, (7)–(10) show that *because*-X lacks any syntactic integration with the main clause, and the lack of negative scope

ambiguity discussed in Section 1 further suggests that *because*-X clauses are syntactically 'inert'.

Secondly, X does not seem to 'stand in for' propositional content — unlike standard *because*, the intended meaning of a *because*-X sentence cannot be 'translated' into a fully lexicalised, non-reduced form. Rather, *because*-X is a device for conveying potentially an array of related meanings. Consider the constructed examples in (28) and (29).

(28)   Sibling 1: Is Vic coming with you for Thanksgiving?
         Sibling 2: Probably not, because mom.

(29)   A: Is everyone pumped and ready to go for Clearing?
         A′: Are you going to be at the meeting tomorrow?
         A″: Why do I already kind of think this after-work 'drinks' will be lame?
         B: Yeah, because ☕

If forcibly translated into propositional form, *because mom* in (28) could encompass all of 'because mom is sensitive about this topic', 'because mom has a stubborn personality', 'because mom has high blood pressure', and 'because I don't want to fight with mom'. Similarly, *because* ☕ in (29) can be a coherent response to any number of versions of A because its out-of-context meaning is not fixed: even in response to A, it can mean 'I'm ready for Clearing because I've had a huge number of cups of coffee this morning', or 'I'm ready for Clearing because I've had my dependable single cup of coffee as I do every morning'. As a response to A′, B could mean 'I'm going to the meeting because there will be free coffee there', or 'I'm going because the meeting is in my favourite coffeeshop and I'll take any excuse to go and get a delicious cortado'. In response to A″, B could mean 'I agree the after-work drinks will be lame because we will be expected to order coffee, as opposed to alcoholic drinks', or 'It'll be lame because the organiser is stingy and will probably not pay for alcoholic drinks, only coffee.' In each of these cases, the context of use fills in a great deal of what is conveyed.

5.2 What *because*-X is (at the moment)

As (28) and (29) demonstrate, the interpretation of *because*-X relies more heavily on the common ground (Stalnaker, 2002) shared by a speaker and an addressee, and in fact, the same *because*-X can receive a different interpretation depending on what background information is shared. In this respect, *because*-X is like react emoji use on social media: a ❤️ react can be used to express shared happiness

with someone posting good news, shared sadness with someone sharing sad news, or even to indicate that a post is heartwarming, or relatable. The heavy dependence on the common ground makes the conditions of use for *because*-X much narrower than for standard *because*.

A feature of attested uses of *because*-X is that X generally represents a state (emotional, mental, societal or physical), or state of affairs. This is suggested by the contrast between the progressive forms in (30-a) and the past tense counterparts in (30-b). Note that, as long as what is expressed is an emotional state (30-c), X need not be constrained to single words.

(30) a. because {falling asleep, waking up, running late}
      b. *because {fell asleep, woke up, ran late}
      c. because {wtf, <groan>, why is this happening to me, how the hell could you say that, damnit, holy shit}
      d. because {computer, Obama}

Even canonical cases involving bare nouns or proper names, like (30-d), must be interpreted to have an overarching meaning like 'because of the state of my computer right now', or 'because of the situation with Obama', whatever the specific situation might be (dependent on the speaker's political views).

We suggest that, at present, there are two related subtypes of *because*-X which have slightly different characteristics. Many cases of *because*-NOUN exemplify the deictic subtype, whose function is to gesture at some real world referent or situation. These uses (e.g., (30-a), (30-d)) appear to resist multi-word expressions in X, and need not have a particular expressive meaning: *because* ☕ or *because coffee* need not be interpreted sarcastically, for example.

By contrast, the expressive subtype functions specifically to convey emotional content. As demonstrated by (30-c), X can easily be a multi-word expression, and superficially conforms to grammatical rules: the X in *because seriously what is going on right now?* is superficially a well-formed *wh-* question. What is conveyed by expressive *because*-X in terms of information content (31-a) is no different from what is conveyed by its standard *because* counterparts (31-b).

(31) a. because {millenial, artist, PhD student}
      b. because {I am a millenial, I am an artist, I am a PhD student}

This use of *because*-X is comparable to other expressive content such as expressive attributive adjectives (*the damn dog*) or epithets (*that pillock*), where the truth-conditional content (the causal relationship, in the case of *because*) is equivalent but the expressive aspect of the meaning, namely the speaker's attitude or mental state, takes wider scope and the speaker would feel they were

being misrepresented (their full meaning not conveyed, or attitude ascribed to them that they did not hold) if one were substituted for the other in reported speech. As Potts (2007) sets out for other types of expressive content, the expressive meaning of *because* is independent of the causal meaning and states something about the utterance situation or speaker's perspective. They cannot be paraphrased (e.g., with a standard *because* form) without losing the expressive content, and they achieve a communicative goal by virtue of being uttered.

The added meaning conveyed by the examples in (31-a) is both expressive, and dependent on specific shared knowledge, such as what it is like to be a PhD student. Note that these uses display the interpretive flexibility described above, as illustrated in (32).

(32)   a.   A: Are you gonna buy the new XBox?
              B: No, because PhD student (∼No, because I'm poor)
        b.   A: Should I ask him out?
              B: No, because PhD student (∼No, because he's socially awkward/never does things for fun anyway)
        c.   A: Should I ask her how work is going?
              B: No, because PhD student (∼No, because work is undoubtedly frustrating and miserable for her)
        d.   A: Should I ask her how work is going?
              B: Idk, because PhD student (∼I don't know, because you never know with PhD students: sometimes they're really enthused and can't stop talking about their projects, and other times they're frustrated and miserable)

Expressive *because*-X is exemplified by emoji, facial expressions or vocalisations whose function is to convey an emotional state.

The combination of narrow conditions of use and fluid (context-dependent and non-propositional) meaning explains why *because*-X is canonically low register, and used between specific in-group members: such uses require the addressee to share some contextual knowledge (see, e.g., Bell (1984) on audience design and Eckert and McConnell-Ginet (1992) on language in communities of practice). The sarcasm in (19) is only conveyed if the addressee knows that 'SpongeBob' text is used to convey sarcasm, and the meaning of *because mom* in (28) is only conveyed if the hearer knows something about the speaker's family situation, for example.

This paper provides a snapshot of the usage of an emerging construction which is unstable by definition: judgements and usage patterns are highly variable across the language community, though individual speakers are consistent

in their own usage. As the construction continues to develop and spread, we highlight two particular aspects of *because*-X that vary across individuals and may shift over time: first, the two subtypes (deictic and expressive, which already overlap) may become more similar to each other, or diverge and become more distinct. Secondly, a phrasal or multi-word element as X may become more acceptable as *because*-X becomes more broadly available, or a rigidly single-word X may solidify as the only acceptable form.

Bibliography

Michael Adams. Because much? *American Speech*, 92(4):533–546, 2017.

Laura R. Bailey. Because reasons. *LinguistLaura*, 2 July 2012, July 2012. URL https://linguistlaura.blogspot.com/2012/07/because-reasons.html.

Laura R. Bailey and Olivia Seyerle. I didn't write this talk because syntax: A syntactic analysis of 'because NOUN'. Paper presented at *European Dialect Syntax IX*, 23 March 2019, University of Glasgow, March 2019.

Dale J. Barr, Roger Levy, Christoph Scheepers, and Harry J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278, 2013. doi:10.1016/j.jml.2012.11.001.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 2015. doi:10.18637/jss.v067.i01.

Allan Bell. Language Style as Audience Design. *Language in Society*, 13(2):145–204, 1984.

Alexander Bergs. Because science! Notes on a variable conjunction. In Elena Seoane, Carlos Acuna-Farina, and Ignacio M. Palacios-Martinez, editors, *Subordination in English: Synchronic and diachronic perspectives*, pages 43–59. De Gruyter Mouton, Berlin, 2018. doi:10/1515/9783110583571-003.

Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. *Longman grammar of spoken and written English*. Longman, Harlow & New York, 1999.

Emily Blamire. *A syntactic analysis of because x in English...because linguistics*. Manuscript, University of Toronto, 2016.

Justin Bland, Kenneth Jr. Baclawski, and Matthias Raess. *Because formality*: The conjunction-noun construction in online text corpora. Poster presented at the American Dialect Society Annual Meeting, Washington DC., 2016.

Justin Bland, Kenneth Baclawski Jr, and Matthias Raess. The first decade of because-NP: 2007–2016. In *American Dialect Society Annual Meeting*, Salt Lake City, 2018.

Axel Bohmann. Language change because Twitter? Factors motivating innovative uses of because across the English-speaking Twittersphere. In Lauren Squires, editor, *English in Computer-Mediated Communication*, volume 93 of *Topics in English Linguistics*, pages 149–178. De Gruyter Mouton, Berlin, 2016. doi:10.1515/9783110490817-008.

Kate Burridge. *Cos* —A new discourse marker for Australian English? *Australian Journal of Linguistics*, 34(4):524–548, October 2014. ISSN 0726-8602, 1469-2996. doi:10.1080/07268602.2014.929079.

Stan Carey. 'Because' has become a preposition, because grammar. *Sentence First*, 13 November 2013, November 2013. URL https://stancarey.wordpress.com/2013/11/13/because-has-become-a-preposition-because-grammar/.

Rune H. B. Christensen. *ordinal — Regression Models for Ordinal Data*, 2023. URL https://CRAN.R-project.org/package=ordinal. Version 2023.12-4.

Aida Duma. Une innovation en cours: l'emprunt anglais bicoz. *Studia Universitatis Babes-Bolyai-Philologia*, 59(4):103–114, 2014.

Mercedes Durham. Changing attitudes towards the Welsh English accent: A view from Twitter. In Mercedes Durham and Jonathan Morris, editors, *Sociolinguistics in Wales*, pages 181–205. Palgrave Macmillan, Basingstoke, December 2016. doi:10.1057/978-1-137-52897-1_7.

Penelope Eckert and Sally McConnell-Ginet. Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology*, 2:461–490, 1992.

Catherine Fortin. Some (not all) nonsententials are only a phase. *Lingua*, 117(1):67–94, January 2007. ISSN 00243841. doi:10.1016/j.lingua.2005.09.004.

Megan Garber. English has a new preposition, because internet. *The Atlantic*, 19 November 2013, November 2013. URL https://www.theatlantic.com/technology/archive/2013/11/english-has-a-new-preposition-because-internet/281601/.

Lauren Gawne and Jill Vaughan. I can haz language play: The construction of language and identity in LOLspeak. *Proceedings of the 42nd Australian Linguistic Society Conference – 2011*, pages 97–122, 2011. URL http://hdl.handle.net/1885/9398.

Joan B. Hooper and Sandra A. Thompson. On the Applicability of Root Transformations. *Linguistic Inquiry*, 4(4):465–497, 1973.

R Huddleston and G K Pullum. *The Cambridge grammar of the English language*. Cambridge University Press, Cambridge, 2002.

Masaru Kanetani. On the new usage of because. *Studies in language and literature*, 68:63–79, 2015. URL http://hdl.handle.net/2241/00129328.

Andrew Kehler. Coherence and the Resolution of Ellipsis. *Linguistics and Philosophy*, 23:533–575, 2000.

Martin Konvička. De verborgen complexiteit van want/omdat X*. *Internationale Neerlandistiek*, 57(2):161–183, July 2019. doi:10.5117/IN2019.2.004.KONV.

Martin Konvička. Because X: Now I'm an ellipsis and now I'm not. In *Linguistweets: The first international Twitter conference on linguistics*. Abralin, December 2020. URL https://twitter.com/TeapotLinguist/status/1335365957401907200.

Martin Konvička and Kristin Stöcker. (Non-)Ellipses in Dutch, English, and German: The case of because X. *Nederlandse Taalkunde*, 27(3):333–367, December 2022. doi:10.5117/NEDTAA2022.3.003.KONV.

Marcia C. Linebarger. Negative polarity and grammatical representation. *Linguistics and Philosophy*, 10(3):325–387, 1987. ISSN 01650157. doi:10.1007/BF00584131.

Gretchen McCulloch. Because reasons, 2012. URL https://allthingslinguistic.com/post/26542624054/because-reasons.

Gretchen McCulloch. Why the new 'because' isn't a preposition (but is actually cooler), 2014. URL https://allthingslinguistic.com/post/72252671648/why-the-new-because-isnt-a-preposition-but-is.

Gretchen McCulloch. *Because Internet: Understanding the New Rules of Language*. Riverhead, New York, 2019.

Octopus Data Inc. Octoparse Web Scraping Tool & Free Web Crawlers, 2023. URL https://www.octoparse.com/.

Christopher Potts. The expressive dimension. *Theoretical Linguistics*, 33(2):165–198, 2007. doi:10.1515/TL.2007.011.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL https://www.R-project.org/.

Anneliise Rehn. Because Meaning: Language change through iconicity in internet speak. *SURF conference proceedings 2014*, 2015.

Tyler Schnoebelen. Innovating because innovation. *Corpus Linguistics*, 15 January 2014, January 2014. URL https://corplinguistics.wordpress.com/2014/01/15/innovating-because-innovation/.

Robert Stalnaker. Common ground. *Linguistics and Philosophy*, 25:701–721, 2002.

Rachael Tatman. Why you can, can't & shouldn't do with social media data. Paper presented at the Joint Statistical Meetings, 28 July 2018, 2018.

Sandra Thompson, Robert Longacre, and Shin Hwang. Adverbial clauses. In *Language typology and syntactic description*, volume 2, pages 237–300. 2nd edition, October 2007. doi:10.1017/CBO9780511619434.005.

Benjamin Zimmer, Jane Solomon, and Charles E. Carson. Among the new words. *American Speech*, 89(1):89–110, February 2014. doi:10.1215/00031283-2726413.

Appendix A

Materials for exploratory acceptability survey (Section 2.1). Items 1, 2, 6–9, 15–19 are attested examples cited in Carey (2013); items 3–5, 11, 13, 14, 20–22 are constructed examples; item 10 is a modified album title; item 12 is cited in Garber (2013).

Standard *because*-CLAUSE and *because-of*

1. Starting a tumblr because all the cool kids are doing it
2. I was late because I forgot to set my alarm
3. School was closed because of snow

Clause with overt subordinator

4. I was late because that I got lost
5. I'm edgy because if I left the oven on

Interjection

6. That feeling you get when you finish an essay and you just want to cry because yay
7. I can't believe she did that because honestly

Bare noun

8. Studying because school.
9. Dow closes at record high for 35th time this year because Obama

Noun phrase

10. I'm here because the internet
11. My computer is broken because a virus

12. I can't come out tonight because too much homework.
13. I'm going to fail this test because hard questions
14. Going to bake cookies because this really crappy hangover from last night

Adjective

15. Going to bed way early because exhausted
16. A lot of people should really start thinking before they talk because stupid.

Verb phrase

17. stomach ache because laughing
18. Bye going to study for English because didn't finish this morning because fell asleep
19. Those moments when you choose to eat a salad not because you want salad... but because want croutons.
20. Going to draw this character now because kicks ass!

Prepositional phrase

21. I'm so tired because up all night
22. I'll be late because at the doctor's

Appendix B

Materials for acceptability survey (Section 4). Items 1–19 are attested examples from Twitter, and items 20–21 are modified examples. Items 22–30 are constructed.

Interjection

1. i had to crop myself out because ew
2. i had to share this photo because damn
3. an hour of cute kitten content because yes
4. i saw this on instagram and had to post it because wow
5. i doubt the internet can ever be a truly happy space for anyone because whaaaat
6. what is he thinking because honestly

Bare noun

7. i have too many projects on the go because yarn

8. this is a long thread because boredom
9. they have to employ him because balance
10. they wouldn't ask their sons to do that because misogyny
11. i'm going to bed because self-care
12. you can't work in here because covid

Determiner phrase

13. you have to listen to this song because the lyrics
14. the quality is really bad because my computer
15. they must be related because this resemblance
16. i love this photo because the quality
17. if a person doesn't like me it's because their own reasons
18. i can't stop laughing because this pic

Prepositional phrase

19. we should club together and buy gym equipment because at this point
20. excuse my typos because on mobile
21. i wanted to switch accounts because with a fresh start
22. he looks a bit damp because in the rain
23. i can't buy it because out of stock
24. i've given up on this essay because in the end

Bare verb

25. i have a pretty good vocabulary because read
26. i really need to lose some weight because eat
27. as an army captain you are very loud because shout
28. i feel like you don't want to go because linger
29. i can tell when you want help with your laptop because call
30. i never feel lonely with you because talk