

Peter Uhrig*

Hand gestures with verbs of throwing: Collostructions, style and metaphor

<https://doi.org/10.1515/gcla-2022-0006>

Abstract: This paper is an exploratory corpus-based study of a set of verbs of throwing and their co-occurrence with iconic gestures. It is hypothesized that the (in)formality and the metaphoricity of verbs is related to co-speech gesture frequency, but ultimately, relatively little evidence is found for such relationships. A simpler alternative assumption, i.e. that it is mental simulations that drive co-speech gesture alone, has to be dismissed, too, because the frequency of co-speech gesture is markedly different across verbs, ranging from roughly 41 to about 60 per cent (*lob* vs. *fling*, respectively). One possible explanation might be that, just as with co-occurrences between purely verbal signs, some verbs are conventionally associated with certain types of gestures to a higher degree. With sufficient data, we can quantify these associations, making use of statistics developed for research on collocation and collostruction.

Keywords: co-speech gesture; multimodal communication; Construction Grammar; collostruction

1 Introduction and research questions

The idea for this paper grew out of the anecdotal observation that, at least intuitively, the verb *chuck* seemed to occur with a relatively high rate of iconic gestures in the sense of McNeill:

An iconic gesture is one that in form and manner of execution exhibits a meaning relevant to the simultaneously expressed linguistic meaning. Iconic gestures have a formal relation to the semantic content of the linguistic unit. (McNeill 1985: 354)

If we assume this is correct, then the question arises whether this is an idiosyncratic property of the verb *chuck* or whether there are general processes at work.

* **Corresponding author: Peter Uhrig**, Friedrich-Alexander-Universität Erlangen-Nürnberg, peter.uhrig@fau.de

In order to find out, this paper studies the co-occurrence of gestures with various verbs describing acts of throwing.

One question is whether the fact that *chuck* is a markedly informal verb contributes to its perceived propensity to attract gestures. We would thus hypothesize that verbs that are more formal will show a lower co-frequency with (possibly iconic) gestures than verbs that are more informal.

Furthermore, most uses seen for *chuck* seemed to be literal uses, whereas there was a higher proportion of metaphorical uses with *throw*, so our second hypothesis is that metaphorical uses are less likely to be accompanied by (possibly iconic) gestures because the processes may not be simulated mentally in the same way as literal uses (following the Gesture as Simulated Action framework – see next section).

In addition, the factor of verb frequency may play a role, but a hypothesis could be formed in either direction: One could expect that more frequent words will be more expected and therefore will need less gestural support than unexpected ones. On the other hand, if we keep the rate of information relatively constant (as suggested by Coupé et al. [2019] for spoken language), we would expect adding gestures to high-frequency and thus faster-to-process verbs to be more likely than adding them to low-frequency words that already contain a higher amount of information.

This study is exploratory in nature, rather than strictly hypothesis-testing. Even though some inferential statistics are used to test our hypotheses, it must be borne in mind that these hypotheses were developed on the basis of the same dataset (viz. the NewsScape 2016 English corpus).

2 Theoretical background

Generally, if we follow the predictions of the Gesture as Simulated Action (GSA) framework (Hostetter and Alibali 2008, 2019), we would expect that the verbs in question co-occur with gestures with a similar frequency. They all describe an action performed with the hands, thus can be simulated and imagined perfectly. Accordingly, speakers should gesture with a higher rate than for verbs where that is not the case. However, it is questionable to what extent speakers will perform such simulations in the context of metaphorical uses, which themselves may be highly entrenched. Thus if we *pitch a script to someone*, the process of throwing may not even be conceptualized by the speaker. If we *hurl abuse at someone*, then the hands are most likely not involved, which would make a gesture less likely according to the GSA.

According to Kita et al. (2007) and Argyriou et al. (2017), metaphor is predominantly a right-hemispheric phenomenon and thus leads to a preference of left-handed gestures in right-handed subjects. Given that we do not know who is left- or right-handed, we will for now not pursue this question any further.¹ However, we need to be aware that the motivation behind gestures might also stem from the metaphoricity of the verb uses, which would in turn predict a higher level of gesture, but possibly with a preference for the left hand.

While we will be looking at all gestures – not least because the distinctions are often hard to draw, our focus will be on iconic gestures as defined in the introduction. No distinction will be made between iconic and metaphoric gestures simply because it would be impossible to maintain in practice: Distinguishing between an iconic throwing gesture and a metaphorical one is just not feasible or will produce so many borderline cases that the classification would become meaningless. Often, two further types of gestures are recognized, viz. deictic gestures, i.e. pointing gestures, and beat gestures, i.e. gestures that can be used to accentuate, are often rhythmically related to the speech and, as McNeill (1985: 359) states it, “have no propositional content of their own”.

In a constructionist framework, the frequency of co-occurrence is a significant predictor of the construction status of co-occurring items. Thus, as Goldberg puts it:

Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency. (Goldberg 2006: 5)

Thus, if we allow for multimodality in the constructicon (see Uhrig [2021] for a more detailed discussion as well as Hampe et al. [in prep.] and the special issue of *Linguistics Vanguard* edited by Bergs and Zima [2017] for various further theoretical aspects), we could state that a pattern that consists of a particular verb and a corresponding iconic gesture that occur together with sufficient frequency form a multimodal construction. In a naïve way of thinking, we would then be able to determine the construction status of individual combinations, presuming that we can operationalize “sufficient frequency”. But not only is this unlikely to happen since no threshold has been proposed for construction status, it is also

¹ With a large enough corpus and a reliable gesture annotation, an automatic analysis of gesture handedness might be possible if we find a difference large enough so that the proportion of left-handers (around 10 % of the population), potentially distorting the statistics, becomes irrelevant.

cognitively highly implausible that there is such a sharp dividing-line between constructions and other patterns, such that a specific co-frequency needs to be reached before we can speak of a construction.

Furthermore, the question is not only how high a sufficient frequency would have to be, it is also what should be treated as co-occurrences with sufficient frequency conceptually, a problem that has been discussed in linguistic research on collocations for a long time. Let us consider a simple collocation example, viz. the idiomatic binominal *helter skelter*: The word *skelter* occurs 8 times in the 100-million-word British National Corpus (BNC),² 7 times directly preceded by the word *helter*.³ *Helter* exclusively occurs followed by *skelter*. In addition, they occur as the hyphenated form *helter-skelter* 22 times.⁴ Thus even if we count the hyphenated form, we have a total of 29 occurrences of *helter* followed by *skelter*. By contrast, *only girl* (31), *every girl* (33) and *one girl* (135) should be better candidates for construction status if raw frequency were the only criterion. However, it is not at all clear whether these are entrenched as some kind of unit, given their semantic compositionality and the still rather low frequencies, whereas it seems to be quite clear that *helter skelter* is. This problem is usually resolved by using association measures (see Evert 2005, 2008; Evert et al. 2017; Pecina 2005), which go beyond the co-occurrence of two items, also taking into account the individual corpus frequencies of the words in question and thus the probability of finding such occurrences by coincidence. Due to the high individual frequencies of the co-occurring items, the combinations with *girl* listed above would usually be less strongly associated than, say, *helter skelter*.⁵ We would then expect that lexical combinations that are strongly associated, i.e. collocations, are more likely to be stored, which is inspired by research on foreign language teaching and foreign language lexicography (e.g. Sinclair 1991), where the choice of appropriate collocations is a problem even advanced learners seem to struggle with (see e.g. Nesselhauf [2005] for an overview).

² The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/>

³ Note that one of these 7 tokens is the misspelling *healter*, marked with (sic!), that was used in a report of the LaBianca Murders by Patricia Krenwinkel of the Manson Family. Overall, the combination is predominantly capitalized when used as separate words (most of the time referencing the Beatles' album) and in lower case when hyphenated.

⁴ One of the 7 *helters* and *skelters* that occur as separate words is most likely also a hyphenated form that was wrongly tokenized during corpus creation due to a long dash intervening between them instead of a short hyphen.

⁵ The exact ranking may be very different according to the association measures used; see the references given above for details.

With the notion of collocations (Gries and Stefanowitsch 2004a, 2004b; Stefanowitsch and Gries 2003, 2005), the concept of co-occurrence and association was applied to larger units in Construction Grammar, in particular to determine the most strongly associated (and thus often most prototypical) lexical fillers of grammatical constructions. As opposed to traditional collocation research, Stefanowitsch and Gries (2003) did not only calculate which items are strongly associated ('attracted'), but also which items occur less frequently together than expected, complementing the notion of "attraction" by that of 'repulsion'.⁶ In Uhrig (2021), the notion of collocation is generalized to constructions of arbitrary sizes on arbitrary levels of representation that co-occur together more frequently than expected, based on Proisl's (2019) generalized model for the co-occurrence of linguistic structures. The concept of 'crossmodal collocation' is then introduced as a means to quantify the degree to which, say, a word and a gesture are associated. Again, association measures are calculated in order to determine this degree of association and to be able to rank candidates on a cline that ranges from free combinations (no association at all) via moderately associated items (crossmodal collocations) to items that are so strongly associated that they very often occur together, which would make them prime candidates for multimodal constructions.

For this paper, we are not interested in the extreme ends of the scale. We expect all our verb-gesture combinations to be in the crossmodal collocation zone, but we may be able to see differences in the respective degrees of association.

3 Methodology

First, a list of (near-)synonyms of *throw* was compiled, which included the lemmas *cast*, *catapult*, *chuck*, *fire*, *fling*, *hurl*, *launch*, *lob*, *sling*, *pitch* and *toss*. These were then searched for in the NewsScape 2016 gesture-annotated corpus with commercials removed (to the extent possible), totalling 234 million tokens (see Uhrig [2018] and [2021] for details on the corpus). The second column in Table 1 gives the raw frequencies of the verb lemmas in descending order.

Three very frequent verbs (i.e. more than 10,000 hits in our corpus) had to be excluded from the study because their polysemy patterns would have made it too

⁶ Whether or not actual instances are found in the corpus, 'repelled' items are furthest away from the strongly associated combinations on the cline.

Table 1: Lemma frequencies for the verb *throw* and its synonyms without and with an automatically detected person on screen; manually-annotated sample in last column.

Verb	Lemma frequency	Token No: person on screen	Sample annotated
<i>throw</i>	34,117	23,598	200
<i>toss</i>	3,745	2,262	200
<i>pitch</i>	1,926	1,234	200
<i>hurl</i>	606	426	200
<i>catapult</i>	306	(215)	306
<i>lob</i>	295	(206)	295
<i>sling</i>	188	(131)	188
<i>fling</i>	177	(101)	177
<i>chuck</i>	53	(32)	53

difficult to find a sufficient number of tokens relevant to our research question: The verbs *cast* and *launch* were immediately excluded because they were predominantly used in contexts such as casting actors for a role or launching a new film or product. The verb *fire* was mainly used meaning ‘terminate employment’ and was thus removed, too.

Some of these verbs were still too frequent to annotate the entire dataset manually for the presence or absence of gesture, so that a further step of creating a sample by random thinning was applied. Since a large number of instances have to be discarded in the manual analysis of co-speech gesture in multimodal news corpora, because the speaker is not visible, the hands of the speaker are not visible, or the transcript is wrong, it was decided to increase the odds of finding usable corpus hits by applying a computer vision filter (Turchyn et al. 2018; see Uhrig [2021] for a discussion and evaluation) which requires that a person be detected on screen. While this does not protect against wrong transcripts, camera angles without hands, or voiceovers, it can reduce the number of stock charts, weather maps, live coverage of hurricanes or car chases, etc. one would have to click through and discard. Using this approach is only possible, though, when there are so many hits in the corpus that some have to be discarded, as the software used has a very good precision (i.e. if it says there is a person on screen, then there is one in more than 95 per cent of the cases) but a much worse recall (i.e. it will fail to recognize people in a considerable number of cases).⁷

All verbs with up to roughly 300 hits (i.e. up to *catapult* in Table 1) were annotated in toto. The concordances for the more frequent verbs were first filtered with

⁷ Exact recall/precision figures are currently not available for this system.

the computer vision system, which removed between 30 and 40 percent of the data. The remaining hits were then randomly thinned to 200 in CQPweb (Hardie 2012). The concordances were downloaded and uploaded into the Rapid Annotator, a web application specifically designed for the relatively fast classification of a large number of items (text, image, audio, video; see Uhrig 2021 for details).⁸ For each verb, a separate experiment was created, always with the configuration presented in Table 2:

Table 2: Configuration of questions and possible answers in the Rapid Annotator.

<i>1. Is this a literal or a metaphorical use?</i>		
Options	– literal	
	– literal but not human scale/agent	
	– metaphorical	
	– undecided	
	– problem	→ skip remaining questions
	– repetition	→ skip remaining questions
<i>2. Is there a gesture?</i>		
Options	– yes	
	– no	→ skip remaining questions
	– no speaker/hands visible	→ skip remaining questions
	– problem	→ skip remaining questions
<i>3. Type of gesture?</i>		
Options	– iconic	
	– deictic	
	– beat	
	– unclear/hard to see	

It has to made very clear that the simplifications made here are (and have to be) to some extent unsatisfactory. Thus there are good points to be made for assuming degrees of metaphoricity instead of a binary opposition (see e.g. Müller 2008: Ch. 6) and, as we will see below, the different types of gesture are not mutually exclusive.⁹ There will be borderline cases and instances where other annotators might have decided differently, but a total of 1,819 videos were annotated within

⁸ See <https://github.com/RedHenLab/RapidAnnotator-2.0/>.
⁹ In the meantime, the option of selecting multiple categories has been added to the Rapid Annotator to be able to account for such phenomena without spelling out all possible combinations.

this schema and another 200 for metaphoricity (see below) so that such cases should not change the overall trend of the results.

The biometric clustering function of Rapid Annotator was used in order to cluster repeated instances together so as to be able to remove the repetitions, which otherwise can be hard to spot and would have skewed the results.

Concerning the classification of a corpus hit as literal or metaphorical, samples with a person on screen may exhibit distributions which are different from those in unfiltered samples. Thus, in order to obtain numbers comparable to those of the verbs for which no sampling had taken place, four completely random samples of 50 hits each for *throw*, *toss*, *pitch* and *hurl* were uploaded to separate experiments in the Rapid Annotator. These were set up to only contain the first question with the same options as given in Table 2 above.

In order to determine the colloquiality or informality of the verbs, existing learners' dictionaries were consulted – the *Longman Dictionary of Contemporary English* and the *Macmillan Dictionary* – to check for labels such as *formal*.¹⁰ Unfortunately, the dictionaries only agree on the overall informality of *chuck* and some uses of *lob*. Moreover, a simple informal/non-informal dichotomy is most likely too coarse. For the purpose of this paper, a very simple informality ratio was thus calculated on the basis of the British National Corpus (BNC). The original idea was to divide the relative frequency of the verb in question in the most informal subcorpus available by its relative frequency in the most formal subcorpus available and take the logarithm (ln) of this ratio to obtain a symmetrical value. The spoken demographically-sampled component is certainly the most informal one in the BNC; but for the most formal one, the choice is much more difficult. While the written academic portion is certainly quite formal, it differs from spoken everyday language along so many dimensions (mode, situatedness, addressees, ...) that it would be quite unclear what exactly this ratio measures.

To resolve this, we decided to measure informality separately from “spokenness” and introduce two types of ratio: The *informality ratio* is now defined as the logarithm of the relative frequency in the spoken demographically-sampled subcorpus divided by the relative frequency in the spoken context-governed subcorpus. A small offset of 0.001 was added both to the numerator and the denominator, so that (a) no division by zero problems occurred and (b) no illegal logarithms were created. The *spokenness ratio* was calculated with the entire spoken and written subcorpora. Thus, it is again the logarithm (ln) of the relative frequency in spoken texts divided by the relative frequency in written texts,

¹⁰ See <https://www.ldoceonline.com/> and <https://www.macmillandictionary.com/>, respectively.

this time without offset because there were no zeroes to compensate. The frequencies in the spoken subcorpora as well as the ratios are reported in Table 3. Although we can calculate moderate correlation coefficients ($r = 0.54$, $p = 0.13$; $\rho = 0.52$, $p = 0.16$), these are not significant and a correlation between the two ratios can neither be confirmed nor denied, although we had expected a moderately positive correlation because, overall, spoken language is less formal than written language.

Table 3: Verb lemma frequencies in spoken subcorpora of the BNC with calculated informality and spokenness ratios.

Lemma	Spoken Demographic		Spoken Context-Governed		Informality Ratio	Spokenness Ratio
	Raw freq.	Rel. freq. p.m.w.	Raw freq.	Rel. freq. p.m.w.		
<i>chuck</i>	248	58.57	55	8.91	1.883	1.931
<i>fling</i>	21	4.96	9	1.46	1.222	−1.578
<i>sling</i>	28	6.61	15	2.43	1.000	0.078
<i>throw</i>	653	154.23	505	81.77	0.635	0.001
<i>toss</i>	25	5.90	24	3.89	0.416	−1.032
<i>lob</i>	7	1.65	8	1.30	0.238	0.102
<i>pitch</i>	9	2.13	33	5.34	−0.919	−1.188
<i>hurl</i>	2	0.47	8	1.3	−1.016	−2.070
<i>catapult</i>	0	0	2	0.32	−5.771	−1.555

Obviously, the advantage of the spokenness ratio is that, due to the larger subcorpus sizes, no zero frequencies are involved, which is why such extreme cases as in the informality ratio (here *catapult*) are attenuated. Still, since the spoken component of the BNC contains data collected in relatively formal settings and since the written part contains genres that mimic spoken interaction, it does make sense to keep both measures and not to simply use the spokenness ratio as a proxy for informality just because it is based on more data.

4 Results

First of all, let us look at gesture frequency in our dataset by verb, as given in Table 4. The rate of gesture use in our dataset is between 41 and 66 per cent. This rate is calculated on the basis of only those cases in which a gesture was either visible or not visible, but excluding those where no speaker or no hands were visible or which were skipped for other reasons.

Table 4: Gesture use with the various verb lemmas.¹¹

Verb lemma	Skipped/ problem	No speaker/ hands visible	Gesture absent	Gesture present	Percentage gesture (over yes/no)
<i>fling</i>	57	70	17	33	66.00%
<i>chuck</i>	17	19	6	11	64.71%
<i>pitch</i>	29	74	37	61	62.24%
<i>catapult</i>	60	133	43	70	61.95%
<i>throw</i>	22	60	54	64	54.24%
<i>hurl</i>	33	57	58	52	47.27%
<i>toss</i>	72	46	45	36	44.44%
<i>sling</i>	37	73	45	33	42.31%
<i>lob</i>	67	120	63	45	41.67%
TOTAL	394	652	368	405	52.39%
AVERAGE					53.87%

We can use Table 4 as the basis of a set of calculations of association measures, the results of which are reported in Table 5.

Table 5: Crossmodal collostructional analysis of verb and gesture use (only *Gesture present* and *Gesture absent*) with extrapolated numbers, following the model in Uhrig (2021).

Verb lemma	Cofreq with gesture	Cofreq with gesture (extrapolated)	Collostructional strength	Odds ratio
<i>fling</i>	33	33	1.262	1.669
<i>chuck</i>	11	11	0.591	1.575
<i>pitch</i>	61	587	7.275	1.437
<i>catapult</i>	70	70	1.310	1.401
<i>throw</i>	64	10917	3.455	1.139
<i>hurl</i>	52	158	2.012	0.768
<i>toss</i>	36	674	13.585	0.669
<i>sling</i>	33	33	1.564	0.629
<i>lob</i>	45	45	2.127	0.612

The collostructional analysis performed here classified all occurrences of verbs with a speaker sufficiently visible. This amounted to a total of 773 occurrences, i.e. the sum of the columns “Gesture present” and “Gesture absent” in Table 4. In

11 Note that the value given as the “Total” under “Percentage gesture” is the percentage over the totals of the columns “Gesture absent” and “Gesture present” and not an average of the percentages, giving more weight to the verbs with larger sample sizes.

order to be able to run a collostructional analysis, the cofrequencies determined manually on smaller samples were extrapolated to the full dataset of the verbs of throwing studied here.

For this type of dataset, a traditional approach using the Fisher Exact Test as proposed by Stefanowitsch and Gries (2003) is problematic because the Fisher Exact Test is a test of significance and accordingly shows a bias in favour of higher frequency, because there is more evidence to support the results. Consequently, the verb with the lowest co-frequency in our study, *chuck*, is ranked last in terms of collostructional strength (which is the negative $\log_{10}$ of the p -value), and the first three are the verbs with the highest co-frequency with gesture.

In our example, this is of course particularly problematic because the numbers for these three verbs are extrapolated and accordingly the absolute p -values are meaningless. They have been omitted from Table 5. The ranking by collostructional strength thus illustrates why, time and again, odds ratio, a measure of effect size, has been proposed to be used to quantify and rank the attraction/repulsion of collo-items instead of measures of significance (see the discussion in Evert [2008] or Schmid and Küchenhoff [2013] for advantages and disadvantages). In our simple case of gesture vs. no gesture, the ranking offered by odds ratio is identical to the ranking by percentage in Table 4. With this kind of dataset (i.e. the small sample size), one also has to be aware that the column “Attracted/Repelled”, which was introduced to such tables by Stefanowitsch and Gries (2003), does not make sense because anything that occurs more frequently with a gesture than the average verb in the dataset does will be counted as attracted, and all others as repelled. This says absolutely nothing about whether gesture use is attracted or repelled by those verbs in the corpus in general and would thus be quite misleading. Accordingly, the column has been omitted in Table 5.

Overall, we see that of the 1,819 video snippets we started with, only 404 (roughly 22 per cent), were found to contain the verb in question and a gesture.¹² Of these, another 121 were not sufficiently visible to attempt a classification and discarded from further analysis. The video snippet behind the QR code on the right (click or scan it to watch the video) illustrates this. We can see the body move when the speaker uses the verb *chuck*, and a small part of the hand briefly pops up behind the so-called lower third. If one had to hazard a guess, one would probably classify it as an iconic chucking gesture. Thus, we ended up with a total of 283 classified gestures, corresponding to only 15.6 per cent of the full dataset. Table 6 reports the full set of numbers, and we can see that even though we started with a sizable dataset, we are now often



¹² Due to repeated instances, there are more, but repetitions must be excluded from the statistics where possible, so that the number 404 is appropriate here.

down to single-digit numbers, which make statistics rather unreliable. Thus, even though the percentages given in the last column of Table 6 look very precise, we must be aware that even small changes in some of these lines might have a large effect, i.e. one should not base strong arguments on these numbers.

Table 6: Summary of the analysis of gesture type (percentage of iconic uses calculated only over the usable columns “Beat”, “Deictic” and “Iconic”).

Verb	Skipped	Unclear/ hard to see	Beat	Deictic	Iconic	Percentage iconic	Odds ratio
<i>toss</i>	164	10	0	3	23	88.46%	4.258
<i>chuck</i>	42	2	2	0	7	77.78%	1.795
<i>lob</i>	250	14	5	2	24	77.42%	1.836
<i>fling</i>	144	4	5	2	22	75.86%	1.663
<i>throw</i>	136	15	12	4	33	67.35%	1.051
<i>hurl</i>	148	17	9	4	22	62.86%	0.837
<i>pitch</i>	140	26	11	3	20	58.82%	0.690
<i>catapult</i>	236	25	13	7	25	55.56%	0.576
<i>sling</i>	155	8	12	1	12	48.00%	0.431
TOTAL	1415	121	69	26	188	66.43%	
AVERAGE						68.01%	

Nonetheless, we clearly see that iconic gestures dominate the competition. Although gestures seem to encode both iconic and deictic information in a number of cases, e.g. the action of throwing something and the target direction simultaneously, these were usually counted as instances of the iconic type, so that, with such cases, percentages may be lower for other annotators. The clip behind the QR Code on the right is a typical example, where both the action of throwing but also the direction (*behind her head*) are represented in the gesture.



Let us now return to the hypothesis that informality boosts gesture use, in particular of iconic gestures. As shown in Section 3 above, informality was only determined overall for each lemma because it was deemed impossible to decide on the informality of the context in every video snippet, although this would of course have allowed for more fine-grained and powerful statistics. To test our hypothesis, a correlation test was conducted of (a) the logarithm of the odds ratio for gestures in Table 5 and the informality ratio, and (b) the logarithm of the odds ratio for iconic gestures in Table 6 and the informality ratio. The reason for taking the logarithm of the odds ratio is simply to scale the results the same way as the informality ratio (see discussion in Section 2 above).

In both cases, the *p*-values of the correlation tests are very high (around 0.3 and 0.2, respectively), so that it does not make any sense to report results. One likely explanation is that there is not enough evidence for the nine verbs that we are interested in. Thus, to see whether there is a general tendency for informal verbs to co-occur with gesture, let us create two groups, uniting low-formality and high-formality verbs, respectively: We will put the top 4 of the list into the low-formality (=informal) group and the bottom four in the high-formality group. The verb *throw*, which is right in the middle, will not be considered (it also has a spokenness ratio of practically zero, indicating that it is equally likely to occur in spoken and written language). We can then sum up the results in Table 7.

Table 7: Presence or absence of gesture versus informality or formality of verb.

	Informal	Formal
Gesture	116	209
No gesture	154	150

The figures in Table 7 show that formal verbs show a preference for gesture use whereas informal verbs even disprefer gesture use. The difference is highly significant (*p* < 0.001) and the effect size is considerable (odds ratio = 0.541).

In the next step, we will explore metaphor frequency in our dataset by verb (Table 8).

Table 8: Literal and metaphorical uses.¹³

Verb lemma	Ignored	Literal (all)	Metaphorical	Undecided	% metaphorical	% metaphorical corrected
<i>catapult</i>	60	16	229	1	93.09%	93.09%
<i>throw</i>	21	52	123	4	68.72%	68.29%
<i>hurl</i>	33	41	124	2	74.25%	65.96%
<i>sling</i>	37	52	96	3	63.58%	63.58%
<i>lob</i>	67	82	143	3	62.72%	62.72%
<i>pitch</i>	29	47	121	3	70.76%	48.72%

¹³ Again, the value given as the “Total” under “Percentage gesture” is the percentage over the totals of the columns “Gesture absent” and “Gesture present”, which gives more weight to the verbs with larger sample sizes. By contrast, low-frequency verbs, *chuck* in particular, have equal weight in the “Average”.

Table 8 (continued)

Verb lemma	Ignored	Literal (all)	Metaphorical	Undecided	% metaphorical	% metaphorical corrected
<i>toss</i>	72	44	72	12	56.25%	47.50%
<i>fling</i>	56	76	35	10	28.93%	28.93%
<i>chuck</i>	17	29	6	1	16.67%	16.67%
TOTAL	392	439	949	39	66.50%	–
AVERAGE					59.44%	55.05%

The second column (“Ignored”) provides the sum of problem cases and repetitions. The highest number of these, for *toss*, can be explained by tagging errors produced by Stanford CoreNLP, the language processing system used in the creation of the corpus. Many nominal uses of *toss* in *toss up* and *coin toss* were erroneously tagged as verbs and thus had to be flagged manually. The column “Literal (all)” contains cases identified as “literal” as well as cases categorized as “literal but not human scale/agent”. With only seven examples in the entire dataset, the latter category, which is illustrated in (1), was introduced after realizing that some literal uses were at a scale forbidding any meaningful simulation by the speaker. However, it did not prove to be very useful.

- (1) It’s a tradition that I believe dates back to the Cold War, when it was believed that the fears of the Soviet Union **lobbing** a nuclear missile at the Capitol was possible.
(2016-01-13_0100_US_CNN_Anderson_Cooper_360.txt)

The columns “Metaphorical” and “Undecided” should be self-explanatory. The “% metaphorical” column is the percentage of metaphorical uses among the non-ignored datapoints. In “% metaphorical corrected”, the percentages of the 50-hit random sample were used for *throw*, *hurl*, *pitch* and *toss*, as explained above.

We can clearly see that *catapult* is used metaphorically at by far the highest rate and *chuck* at the lowest. We can also see that it makes sense to look at unrestricted data when determining the percentage of metaphor use, because the samples with a person on screen seem to favour metaphorical uses – all four random samples analysed show a lower rate, still roughly the same for *throw*, but considerably lower for *pitch*, for instance.

The calculation here differs slightly from the one done for our first hypothesis because we have annotations of gesture and metaphoricity on every item in the corpus so that we can look at the relationship between the two without taking the words into consideration. The corresponding 2x2 table is provided as Table 9. The

Fisher Exact Test fails to reach any acceptable significance level so that no direct relationship between gesture frequency and metaphoricity can be attested in our dataset.

To further investigate the part each of them plays in the prediction of gesture use and to find out whether they interact, both aspects discussed as possible predictors of gesture use in this chapter so far can be combined into one generalized linear regression model. Thus a model was created that used not only metaphoricity and the binary formal/informal distinction as predictors, but also their interaction. The model is better than a model without the interaction (lower AIC),¹⁴ and both informality and metaphoricity emerge as significant coefficients ($p = 0.013$ and $p = 0.015$ respectively), but the most significant predictor in the model is the interaction between metaphoricity and formality ($p < 0.001$), which is negative, indicating that informal words in metaphorical use elicit significantly fewer gestures. The model, however, is far from a perfect fit for the data.

Table 9: 2 × 2 contingency table for gesture use with literal and metaphorical verb uses.

	Literal	Not literal (= metaphorical)
Gesture	95	294
No gesture	72	285

5 Discussion

In this study, we were plagued by the same problem that other corpus-based studies of co-speech gesture have been struggling with, i.e. that a very large number of instances have to be screened in order to find usable quantities of co-speech gesture. For instance, in their study of timeline gestures Valenzuela et al. (2020) had to discard 75.34 per cent of the initial hits. In our sample, 57.50 per cent of the initial hits were removed, and some proportion of the difference is certainly due to the computer vision filter looking for persons on the screen.

As expected, *chuck* came near the top of the list of gesture-associated verbs (Tables 4 and 5), but it was beaten by *fling*, illustrated by the example behind the QR code on the left. At the other end of the list, *lob*



¹⁴ It is also better than using spokenness instead or together with informality.

turned out to attract much less gesture use than the average verb in this study. The example behind the QR code on the right illustrates this while also showing that the speaker does not generally shy away from large gestures.

However, we were unable to find a correlation between gesture use and the informality ratio for a given verb. It was only when we grouped the verbs into an informal and formal category based on a positive or negative informality score that we were able to see a significant difference ($p = 0.020$). While this could be attributed to lack of data, we see that both groups are far from being uniform if we plot association with gesture and informality against one another for each verb, as done in Figure 1.

The plot in Figure 1 suggests that, besides the relatively neutral *throw*, which is in the middle concerning the association with gesture, there are two clearly distinct groups of verbs in terms of their attraction to gestures although the collostructional strength indicates that for the small blue dot (*hurl*), there is in fact relatively little evidence for its exact position on the y-axis.¹⁵ The grouping into formal and informal helped us obtain significant results because the top group of four verbs preferring gesture contains two formal ones while the bottom group that disprefers gesture contains one (if at all) in our dataset. Whether this is the best possible explanation of the practically bipartite distribution of gesture association remains doubtful. At any rate, for now we have to report that the opposite of our hypothesis seems to be true, i.e. it is the formal verbs that attract higher proportions of gestures as opposed to the informal ones (at least if we look for a monocausal relationship).

Of course, our approach is much too coarse-grained. After all, many of these verbs can occur in formal and informal settings alike, both in written and in spoken contexts. A future, more in-depth study should thus classify every occurrence in the sample according to the actual (in)formality of its situation of use instead of grouping all uses of one verb together.¹⁶

The plot of gesture association against spokenness ratio, which is presented in Figure 2, presents a similarly hard-to-interpret picture. No sensible claims can be derived from such a distribution.

When it comes to metaphorical uses, *chuck* again behaves as expected, with an extremely low proportion of metaphorical uses, while *catapult* comes out top, which may also be due to the fact that actual catapults are rarely talked

¹⁵ Also, note that the informality ratio for the most extreme point, *catapult*, is an artefact due to lack of data.

¹⁶ It is hard to conceive, however, how this could be objectively operationalized in a TV corpus. There would also be a certain danger that the use of gesture might influence the formality decision, which would be even more problematic.

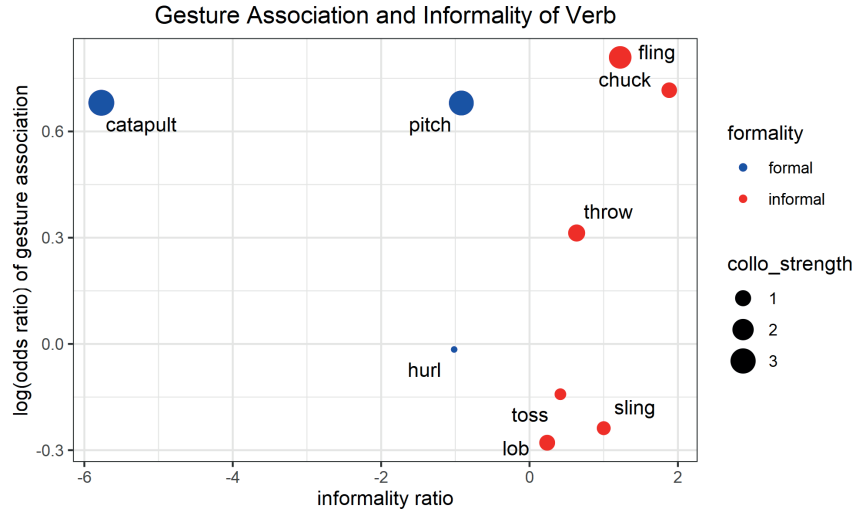


Figure 1: Association with gesture use plotted against informality, grouped by formal/informal status.

about on today’s television. However, the expected direct relationship between metaphoricity and rate of gesture use could not be found, contrary to our original hypothesis.

Although it was not part of our hypotheses, a potential relationship was observed between metaphoricity and informality, in that the strong and significant negative correlation between the two may be taken to indicate that items high in metaphoricity (such as *catapult*) are low in informality and vice versa (e.g. *chuck*) ($r = -0.77$, $p = 0.015$, $\rho = -0.87$, $p < 0.01$). Figure 3 shows the pattern, which is certainly exaggerated by *catapult* in the bottom right corner, but appears to be robust nonetheless. As we should be very careful with any relationships found in the data analysis phase,¹⁷ it is important to stress here again that this study is exploratory and that its results should be regarded as pointers to generating hypotheses for future research rather than as verified facts.

The generalized linear model presented at the end of the previous section is interesting in that it explains the counter-intuitive results we saw in the association between gesture and informality with the help of an interaction. This is to say, it is not generally the case that informality predicts low gesture use, it is the

¹⁷ Starting points for readers not fully aware of these debates may be the search terms “HARK-ing”, “*p*-hacking” and “researcher degrees of freedom”.

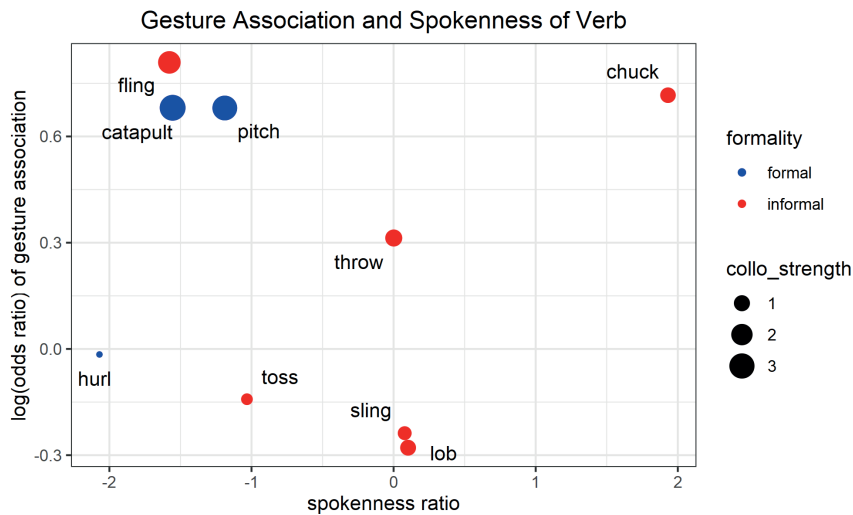


Figure 2: Association with gesture use plotted against spokenness, grouped by formal/informal status.

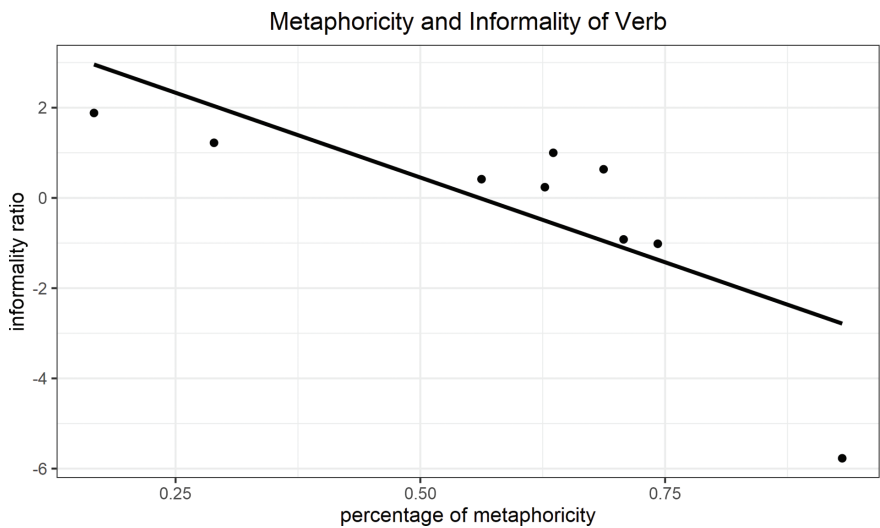


Figure 3: The relationship between metaphoricity and informality.

use of informal words in metaphorical contexts that does. If we account for that, informality actually increases the chances for gesture. But again, given the small number of verbs tested here and the relatively poor model fit, no strong claims should be made based on this model.

6 Conclusion and outlook

In sum, we saw that, while our initial impressionistic notion that the verb *chuck* (1) occurs with gestures quite frequently, (2) is very informal and (3) mostly used in literal contexts all turned out to be true when compared to its near-synonyms, *literalness* does not predict gesture frequency and *informality* does so only under at least slightly dubious groupings into formal and informal verbs. The two predictors may work together, though, but future research is definitely needed to find conclusive evidence for this interaction.

While these negative results may appear disappointing at first, we have to note first that negative results are extremely important for the advancement of scientific research and notoriously underreported (e.g. Fanelli 2012). Moreover, in this case they allow for a different interpretation. After all, what we were looking for in this study was general principles guiding the use of individual verbs. Not finding these is not all that surprising. Research on valency (see e.g. Herbst et al. 2004) and collo-phenomena (see e.g. Sinclair [1991] and the other sources cited in Section 2 above) has time and again shown that item-specific grammatical behaviour and hard-to-predict lexical combinations are extremely common and that native speakers can master these associations by apparently memorizing them together with their conditions of use. Why should this be different for the co-occurrence of verbs with gesture? There may not be a synchronically relevant reason for the differences in gesture use that we see in Tables 4 and 5 other than convention and memory (or such reasons may not account for the full variance found in the data).

There are further aspects to the analysis of the verbs studied in this paper that we were unable to touch upon in detail, even though they would have merited a



discussion in their own right. For instance, as reported above, many individual instances were highly interesting in that they combined multiple gestural functions in one gesture (see Kok et al. [2016] for a discussion). For instance, in the video behind the QR code on the

left, the handshape corresponds to holding the kind of egg-shaped football that is used in American football, representing the object of the verb *throw*, which itself is represented by the throwing motion performed by the speaker. The question of how to model in a constructionist framework the fact that a gesture may represent multiple meanings through multiple form features still needs to be worked out in much more detail. Further open questions are related to whether specific



gestures might have crossed the border between co-speech gesturing and acting (see QR code on the left), and/or how they showed interesting prosodic characteristics, such as the lengthening of the verb (see QR code on the right).



The tools available to researchers now and used in this paper make it easier to tackle datasets of sensible sizes in research on multimodal communication and should thus facilitate further corpus-based research in this area.¹⁸

Acknowledgements

Special thanks go to Philipp Heinrich for his competent help with statistics and R, and to Stephanie Evert, Thomas Proisl and Sebastian Hoffmann for helpful advice. Many thanks also to Beate Hampe, who provided highly valuable feedback on the manuscript. The author gratefully acknowledges the generous funding from the Competence Network for Scientific High Performance Computing in Bavaria (KONWIHR) for the forced alignment of subtitles and audio (project *Robot Hen*) and the automatic person and gesture detection in the corpus (project *Talking Hands*).

References

- Argyriou, Paraskevi, Christine Mohr & Sotaro Kita. 2017. Hand matters: Left-hand gestures enhance metaphor explanation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43(6). 874–886.
- Bergs, Alexander & Elisabeth Zima (eds.). 2017. Towards a multimodal Construction Grammar. [Special Issue]. *Linguistics Vanguard* 3(1).
- Coupé, Christophe, Yoon Mi Oh, Dan Dediu & François Pellegrino. 2019. Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances* 5(9). <https://doi.org/10.1126/sciadv.aaw2594> (accessed 28 July 2022).
- Evert, Stefan. 2005. *The statistics of word cooccurrences. Word pairs and collocations*. Stuttgart: Institut für maschinelle Sprachverarbeitung, Universität Stuttgart, Ph.D. thesis. <http://elib.uni-stuttgart.de/opus/volltexte/2005/2371/> (accessed 28 July 2022).
- Evert, Stefan. 2008. Corpora and collocations. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. An international handbook*, 1212–1248. Berlin & New York: Mouton de Gruyter.
- Evert, Stefan, Peter Uhrig, Sabine Bartsch & Thomas Proisl. 2017. E-VIEW-alation – a large-scale evaluation study of association measures for collocation identification. In Iztok Kosem, Carole Tiberius, Miloš Jakubiček, Jelena Kallas, Simon Krek & Vít Baisa (eds.), *Electronic lexicography in the 21st century: Proceedings of the eLex 2017 conference*, 531–549. Leiden, The Netherlands.

18 For a discussion of the theoretical issues associated with the empirical investigation of multimodal construction grammar, especially using datasets based on TV shows, see Hampe et al. (in prep.).

- Fanelli, Daniele. 2012. Negative results are disappearing from most disciplines and countries. *Scientometrics* 90. 891–904.
- Goldberg, Adele E. 2006. *Constructions at work*. Oxford: Oxford University Press.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004a. Extending collostructional analysis: A corpus-based perspective on “alternations”. *International Journal of Corpus Linguistics* 9(1). 97–129.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004b. Co-varying collexemes in the into-causative. In Michel Achard & Suzanne Kemmer (eds.), *Language, culture, and mind*, 225–236. Stanford, CA: CSLI Publications.
- Hampe, Beate, Irene Mittelberg, Peter Uhrig & Mark Turner. *in prep.* Towards an empirical assessment of the multimodality claim for syntactic constructions. Theoretical and methodological considerations.
- Hardie, Andrew. 2012. CQPweb: Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17(3). 380–409.
- Herbst, Thomas, David Heath, Ian Roe & Dieter Götz. 2004. *A valency dictionary of English*. Berlin & Boston: Mouton de Gruyter.
- Hostetter, Autumn B. & Martha W. Alibali. 2008. Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review* 15(3). 495–514.
- Hostetter, Autumn B. & Martha W. Alibali. 2019. Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin and Review* 26. 721–752.
- Kita, Sotaro, Olivier de Condappa & Christine Mohr. 2007. Metaphor explanation attenuates the right-hand preference for depictive co-speech gestures that imitate actions. *Brain and Language* 101. 185–197.
- Kok, Kasper, Kirsten Bergmann, Alan Cienki & Stefan Kopp. 2016. Mapping out the multifunctionality of speakers’ gestures. *Gesture* 15(1). 37–59.
- McNeill, David. 1985. So you think gestures are nonverbal? *Psychological Review* 92(3). 350–371.
- Müller, Cornelia. 2008. *Metaphors dead and alive, sleeping and waking. A dynamic view*. Chicago & London: University of Chicago Press.
- Nesselhauf, Nadja. 2005. *Collocations in a learner corpus*. Amsterdam & Philadelphia: Benjamins.
- Pecina, Pavel. 2005. An extensive empirical study of collocation extraction methods. In Chris Callison-Burch & Stephen Wan (eds.), *Proceedings of the ACL student research workshop*, 13–18. Ann Arbor, MI: Association for Computational Linguistics.
- Proisl, Thomas. 2019. *The cooccurrence of linguistic structures*. Erlangen: FAU University Press.
- Schmid, Hans-Jörg & Helmut Küchenhoff. 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24(3). 531–577.
- Sinclair, John McH. 1991. *Corpus concordance collocation*. Oxford: Oxford University Press.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2005. Co-varying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1). 1–43.
- Turchyn, Sergiy, Inés Olza Moreno, Cristóbal Pagán Cánovas, Francis Steen, Mark Turner, Javier Valenzuela & Soumya Ray. 2018. Gesture annotation with a visual search engine for multimodal communication research. *Proceedings of the 32nd AAAI Conference on Artificial*

- Intelligence* 32(1). Article 72. Palo Alto, CA: AAAI Press. <https://doi.org/10.1609/aaai.v32i1.11421> (accessed 28 July 2022).
- Uhrig, Peter. 2018. NewsScape and the Distributed Little Red Hen Lab – A digital infrastructure for the large-scale analysis of TV broadcasts. In Anne-Julia Zwierlein, Jochen Petzold, Katharina Böhm & Martin Decker (eds.), *Anglistentag 2017 in Regensburg: Proceedings. Proceedings of the conference of the German association of university teachers of English*, 99–114. Trier: Wissenschaftlicher Verlag Trier.
- Uhrig, Peter. 2021. *Large-scale multimodal corpus linguistics: The big data turn*. Erlangen: Friedrich-Alexander-Universität Erlangen-Nürnberg, Habilitation thesis.
- Valenzuela, Javier, Cristóbal Pagán-Cánovas, Inés Olza & Daniel Alcaraz-Carrión. 2020. Gesturing in the wild: evidence for a lateral, flexible timeline. *Review of Cognitive Linguistics* 18(2). 289–315.