

Bernabeu, P. (2018). *Dutch modality exclusivity norms for 336 properties and 411 concepts* [Unpublished manuscript]. School of Humanities, Tilburg University, the Netherlands. <https://psyarxiv.com/s2c5h>.

Dutch modality exclusivity norms for 336 properties and 411 concepts

Pablo Bernabeu

Tilburg University and Radboud University

pcbernabeu@gmail.com

Part of the toolkit of language researchers is formed of stimuli that have been rated on various dimensions. The current study presents modality exclusivity norms for 336 properties and 411 concepts in Dutch. Forty-two respondents rated the auditory, haptic, and visual strength of these words. Mean scores were then computed, yielding acceptable reliability values. Measures of modality exclusivity and perceptual strength were also computed. Furthermore, the data includes psycholinguistic variables from other corpora, covering length (e.g., number of phonemes), frequency (e.g., contextual diversity), and distinctiveness (e.g., number of orthographic neighbours), along with concreteness and age of acquisition. To test these norms, Lynott and Connell's (2009, 2013) analyses were replicated. First, unimodal, bimodal, and tri-modal words were found. Vision was the most prevalent modality. Vision and touch were relatively related, leaving a more independent auditory modality. Properties were more strongly perceptual than concepts. Last, sound symbolism was investigated using regression, which revealed that auditory strength predicted lexical properties of the words better than the other modalities did, or else with a different direction. All the data and analysis code, including a web application, are available from <https://osf.io/brkjw/>.

1. Introduction

Some words, such as properties and concepts, bear sensory associations which can actually be measured. For example, if you survey speakers about the property ‘blue,’ you will likely conclude that it is strongly visual. In contrast, ‘tangy’ would come out as gustatory or olfactory. These linguistic-sensory associations are used for language experiments. In some cases, these stimuli can be rendered straightforwardly; for instance, if colour words are used (Simmons et al., 2007). In other cases, the creation of the stimuli amounts to studies of their own. It does happen often with an experimental paradigm known as modality-switching, which works as follows. Each trial is made up of a property and a concept word. Participants must verify whether the property can be applied to the concept. For instance, for a trial such as ‘big’ ‘city,’ the correct response would be affirmative. Indeed cities *can* be big. The critical manipulation, covert, is the conceptual modality of trials, and particularly the transition across trials, which can lead to a match or a mismatch of modalities. For example, if the trial |‘big’ ‘city’| appears before |‘sad’ ‘song’|, there would presumably be a mismatch cost. The first trial is primarily visual whereas the second one is auditory. This paradigm was created to test whether modality shifts incur processing costs, where compared to non-shifts. All the rest being equal, this effect would suggest that word comprehension entails sensory processes (Pecher, Zeelenberg, & Barsalou, 2003; also see Pecher, Zeelenberg, & Barsalou, 2004; Solomon & Barsalou, 2004; Newman, Klatzky, Lederman, & Just, 2005; Marques, 2006; Vermeulen, Niedenthal, & Luminet, 2007; Lynott & Connell, 2009; Ambrosi, Kalenine, Blaye, & Bonthoux, 2011; Collins, Pecher, Zeelenberg, & Barsalou, 2011; Hald, Marshall, Janssen, & Garnham, 2011; Scerrati, Lugli, Nicoletti, & Borghi, 2016; Bernabeu, Willems, & Louwerse, 2017).

The modality of the stimuli is absolutely fundamental for such experiments. So how do researchers determine it? They do it by crowdsourcing from the speaker population. They’ll create an experiment or a survey, and gather modality ratings for a large set of words. The resulting corpora are called ‘norms.’ In the area of modality-specific effects (i.e., sensory processes in language), the stimuli used to be created ad-hoc, or otherwise put together from different norms (in addition to the above references, see Barrós-Loscertales et al., 2012; González et al., 2006). Yet, both methods are problematic because they hinder the comparison of effects across experiments. Also, earlier ratings would measure modality in an absolute way, assigning only one dominant modality

to each word. This might not be realistic because conceptual modality might be a continuum rather than a categorical factor. These problems were first addressed in Lynott and Connell's (2009) norms, where hundreds of words received scores on *each* of a series of modalities, not just one. The principal aim was to create new stimuli that would best capture the subtle, sensory load of words. The more fine-grained the norms, the better their performance as experimental stimuli.

2. Conceptual modality

The term 'modality' is broadly used. The type we will focus on here is dubbed *conceptual* modality because it is related to conceptual memories (for all the perceptual associations). In the area of perception, 'perceptual modality' is studied, including its switching effects (Spence, Nicholls, & Driver, 2001; Turatto, Benso, Galfano, & Umiltà, 2002). The term also refers to the mode of presentation for stimuli (Glenberg, 1984), and yet other phenomena in Linguistics (Nuyts, 2001).

Conceptual modality is not a categorical notion. The five modalities analyzed by Lynott and Connell—corresponding to the classical human senses—share the floor with smaller combinations (van Dantzig, Pecher, Zeelenberg, & Barsalou, 2008). Yet other studies have considered modalities such as interoception, exteroception, and proprioception (Ondobaka, Hald, & Bekkering, 2016). Precisely to tackle this heterodoxy, a study has proposed a set of brain-based modalities, which incidentally come up to the dozens (Binder et al., 2016). Yet, there appear to be no necessary modalities for a given experiment. The appropriate modalities depend on the research question.

2.1. Lynott and Connell's norming method

Lynott and Connell (2009) presented respondents with a series of object properties—i.e., adjectives such as 'blaring' or 'blue.' Respondents rated the extent to which they experienced each word through the senses of hearing, olfaction, taste, touch, and vision, with a scale from 0 to 5 for each sense. After averaging the ratings across respondents, each word ended up with scores, or vectors, for auditory, haptic and visual *strength* or *experience* (also described as the 'perceptual strength' in each modality). The highest of those was identified as the *dominant modality*. The three-point vectors were further used to compute the degree of unimodality or multimodality of the terms. This *modality exclusivity* score was achieved by dividing the range of the three modality scores by the sum. The score ranges from 0 to 1, but is often reported as a percentage. The higher

it is, the more unimodal the word; the lower, the more multimodal. Let's consider an example. A word, 'vanilla,' has the following mean scores: Auditory 0, Gustatory 5, Haptic 1, Olfactory 5, Visual 3. Its modality exclusivity would be $5 \div 14 = .36 = 36\%$. The word could be described as mildly multimodal, or bimodal. Modality exclusivity, thus, is a unique index, different from any of its components, or their sum.

In addition to those indices, Connell and Lynott (2012) underscored the importance of general perceptual strength, which is a score equivalent to the score of a word on its dominant modality (that is, the highest score). This index is different from modality exclusivity in that it provides a more robust measure of a word's perceptual content. As Connell and Lynott showed, the overall perceptual strength outperforms abstractness and imageability indices in predicting reading performance.

The property norms were validated in a modality-switching experiment. As the authors expected, their scale-based modality norms enabled a switching cost with an effect size, $d_z = 0.513$, much greater than that of the earlier categorical norms, $d_z = 0.192$.

Later, Lynott and Connell (2013) created new norms for concepts. These, added to the properties norms, led to a series of findings. First, dominantly visual words were by far the most numerous (see also Winter, Perlman, & Majid, 2018). Second, haptic and visual experience were quite related, as were gustatory and olfactory experience. In contrast, auditory experience was relatively independent (see also Winter, 2016). These different relations are important because they may correlate with different levels of detail in word comprehension. Indeed, Louwerse and Connell (2011) found that the distinction between the two pairs of related modalities (haptic-visual and gustatory-olfactory) became less relevant in conceptual processing when participants responded more slowly. The third finding of Lynott and Connell (2013) was that the three modalities presented differences in modality exclusivity, with auditory and visual words showing greater unimodality, and haptic words showing greater multimodality. Fourth, properties were more unimodal than concepts. This, however, may have been influenced by the fact that properties were selected based on modality—leading to more unimodality—whereas concepts were randomly selected (Winter, 2016). Fifth, the modality scores served to demonstrate the sound

symbolism effect, which holds when the sound of a word bears an iconic relation to its meaning (see also Winter, Perlman, Perry, & Lupyan, 2016).

There are alternative norms worthy of attention. Van Dantzig et al. (2011) collected norms for properties and concepts combined. This approach allowed for the control of an important factor left loose in the other norms. Where respondents are asked about properties alone, there is a hair of variance uncontrolled because processing a property likely requires projecting it onto a concept, however unconsciously. Van Dantzig et al. curbed on this, and argued that such a method may be superior in the creation of stimuli for such tasks as property-verification. For the present replication, however, it was deemed more convenient to apply the separate norming method because it would allow for the comparison of properties and concepts.

3. Conceptual replication in Dutch

The current study reproduces the Lynott and Connell norms for properties and concepts, shifting to words in Dutch rather than English (see also Speed & Majid, 2017). For this purpose, the original materials were translated. The target analyses are determined by the findings above described. Particularly, the last four are most important. That is so because we can reproduce them without the confound influence of the translation: The fact that the terms tested are translated from the English norms does not affect the analyses for those findings. While the translation only attended to the meaning and dominant modality of the source term, the last four findings are based on deeper, language-internal relations.

Where unspecified, all English data referred to below refers to Lynott and Connell's (2009, 2013) norms.¹ Also, where unspecified, 'Lynott and Connell' refers to both norms. The purpose of this conceptual replication is twofold: there is a methodological-linguistic aspect, and a psychological one. In the former case, these materials should facilitate the composition of Dutch experimental stimuli, and perhaps non-academic materials too. Second, this reproduction allows us to re-test research findings, which is important for two reasons in turn. First, generalization across languages has been a long-standing concern in the language sciences (Evans & Levinson, 2009;). Just as we cannot assume the consistency of behaviour across or even within cultures, we should be wary too

¹ Materials retrieved from <http://www.lancaster.ac.uk/people/connell/lab/norms.html>.

of cross-linguistic differences, even for languages such as English and Dutch. Accordingly, the two language samples will be compared per se. Second, the reassessment of psychological findings is an important task. During the last decade, this issue has received particular attention in the context of insufficient or failed replications, frail statistical methods, etcetera (Open Science Collaboration, 2015; see also Baker, 2016). Even if the field were better off than that (Gilbert, King, Pettigrew, & Wilson, 2016), replication is important.

The Lynott and Connell norms, along with a reanalysis by Louwse and Connell (2011), reveal an interesting interplay across modalities. Specifically, the visual and the auditory modality were quite related, and so too were the gustatory and olfactory modalities. In contrast, the auditory modality kept to itself. The present study will test for such an interplay, but as thriftily as possible. Three modalities will suffice. The gustatory and olfactory modalities are spared based on how scarce they were in the English norms. This scarcity most likely would hold for Dutch as well, because it corresponds to the natural lexica of Indo-European languages, where those modalities have just enough words (Burenhult & Majid, 2011). For all cross-linguistic comparisons, the English norms were reanalyzed with the three relevant modalities alone.

3.1. Method

3.1.1. Respondents

The modality rating was completed by forty-two university students of Radboud University and Tilburg University, in the Netherlands, who were not paid.

3.1.2. Materials

The tested words were mostly based on the norms of Lynott and Connell (2009, 2013). Particularly, 336 properties and 387 concepts from the auditory, haptic and visual dominant modalities therein were translated into Dutch.² On top of that, twenty-four concepts were added, each of which was

² Besides the olfactory and gustatory words which were not included, twelve words from the relevant modalities (seven properties, five concepts) were removed because their translation became too difficult, and they were dispensable. The actual problem was an overlap with other translations (e.g., ‘gold’ and ‘money,’ both *geld* in Dutch). Surely it would have been possible to translate them, but it seemed more efficient to just create new terms. In any case, some had to be added in order to fulfill the stimuli for the other experiment.

created to potentially adhere to a particular modality (this was done to create stimuli for an experiment, in which modality switch effects were elicited across trials; Bernabeu et al., 2017). The final sum of items in Dutch was 336 properties and 411 concepts. All properties were of the category adjective. Most if not all concepts were of the category noun, the rest being adverbs and verbs (see Supplementary Materials).

The translation was performed separately for properties and concepts. For both alike, the utmost principle, after the pure meaning of the words, was to render terms that would keep the dominant modality of the source term. Thus, the creation of all terms alike was ‘modality-bound.’ This will be relevant when it comes to comparing the results for properties and concepts. The naturalness of the translations was the second criterion. In contrast, modality exclusivity was not a translation criterion: Neither perceptually stronger nor weaker words were sought. Translation shifts were sometimes necessary. For example, the colour ‘tangerine,’ absent as such from the Dutch language, was translated as ‘fuchsia’—with ‘orange’ and ‘red’ already taken.

A series of online language tools were instrumental in the translation. At a first stage, all properties and concepts were entered separately into the Google automatic translator (for its hegemony among similar tools, see van den Bosch, 2008). Then, the following tools ensued: 2006 Collins Cobuild English dictionary (CD-ROM); online English-Dutch dictionary *bab.la*; English thesaurus *thesaurus.com*; Dutch thesauri *woorden.org* and *synoniemen.net*; database for real translation cases, *linguee.com*; Google Search, Google News, and Google Books, where exact strings (“ ”) were looked up within reliable sources; Princeton University’s Wordnet corpus, *wordnetweb.princeton.edu*, with the use of ‘sister term.’ Throughout this process, we were helped by two Tilburg University students, native Dutch speakers.

Although the English terms made a convenient stepping stone, they did not constitute an end in themselves. Thus, the modality scores that bind for the Dutch norms are only those of the Dutch norming. Any possibly mistranslated terms are just as valid, because they were independently tested (e.g., concept ‘die,’ translated as *sterven* ‘perish/perishing,’ but actually meaning a metal cutting cast).

The Dutch items were normed by means of a survey. All the terms were initially split over two lists, one with 336 properties, the other with 411 concepts. Six respondents completed either of

those. From that point, because those lists would take too long to complete (about an hour), each list was split into three more in turn (splitting was also done in both of Lynott and Connell's norms). This was done pseudo-randomly, so that each resulting list contained a comparable number of items from each of the source dominant modalities.

Analysis file. For the cross-linguistic comparison, the Lynott and Connell data from the auditory, haptic, and visual dominant modalities—including the few items that had not been used in the Dutch norms—were appended to the Dutch norms, coupling both norms for most (though not all) terms. All English items from the three relevant modalities were included, namely, 343 properties and 392 concepts. The variables copied were: word, dominant modality, strength of each modality, and modality exclusivity. In the file, the column 'normed' indicated whether a term was normed in either language or in both (and therefore coupled). Note that, for each of the comparisons, there may be an unbridgeable distance between the two norms due to the different number of modalities tested in each. The underlying conceptual space of respondents to the English norms, who were asked about five modalities, could have been wider than for our respondents. Yet, on the positive side, the comparisons between languages will be partly 'normalized' by the three same modalities analyzed in both samples.

Norms file. There is also a reduced version of this file, focused on the Dutch norms, which is named 'norms.csv.' Herein, the English words are not present, even as translations, because they do not make good translations in all cases (see Supplementary Materials).

3.1.3. Procedure

The procedure was similar to that of Lynott and Connell (2009, 2013). Concepts and properties were separately rated. Respondents were asked to rate each word on the auditory, haptic, and visual modalities, leaving any unknown words blank. Unlike the standard experimental setup implemented by Lynott and Connell, the present norms were collected through a survey.³ Respondents completed the survey at home or wherever they chose, over the course of a fortnight.

³ Survey instructions: 'This is a stimulus validation for a future experiment. The task is to rate how much you experience everyday' [properties/concepts] 'using three different perceptual senses: feeling by touch, hearing and seeing. Please rate every word on each of the three senses, from 0 (not experienced at all with that sense) to 5 (experienced greatly with that sense). If you do not know the meaning of a word, leave it blank.'

3.1.4. Design and analysis

In the analysis file, Dutch and English data were described in separate columns mostly (see Supplementary Materials). All analyses were separate for properties and concepts, except for a translation check. The statistics computed were (specifying treatment of English and Dutch norms): reliability analysis (only Dutch norms), Pearson's correlation (norms independent and paired), one-sample t-test (norms independent), Principal Components Analysis (norms independent), ANOVA (norms paired), and multiple regression (norms independent).

3.1.5. Preprocessing

Forty-two surveys were received, one of them completed only up to a half. Due to the split of the surveys in sections (to reduce the time required), and a number of drop-outs (surveys were distributed sequentially), measures were collected in slightly different proportions across words. On average, there were eight raters per word, with a minimum of five and a maximum of ten. The average of blank cases—meaning unknown words—per survey was 1.31%.

In order to attain one single score per word on each modality, the ratings from all respondents needed to be collated, as in Lynott and Connell (2009, 2013). This process, done in Excel, resulted in 1,233 unique data points in the concepts sample, and 1,008 points in the properties sample. The appropriateness of the averages was calculated through reliability analysis (see analysis code and output in Supplementary Materials). Two measures were calculated (Cronbach, 1951). 'Interrater reliability,' λ_6 , measures the fit among raters, independently of items. By contrast, 'interitem consistency,' α , does the opposite, measuring the fit among items independently of raters. For both measures, a conventionally satisfactory minimum is $\alpha = .70$ (Kline, 1999; Field, Miles, & Field, 2012; Woodruff & Wu, 2012). Overall, the interrater reliability was sufficient, with all scales above $\alpha = .70$, and an average $\alpha = .75$. Interitem consistency—also known as 'squared multiple correlation,' or G_6 —was fine too, with all scales above $\lambda_6 = .70$, and an average $\lambda_6 = .79$. The individual figures were as follows. In the case of properties, the Auditory scale had a medium interrater reliability, $\lambda_6 = .78$, and interitem consistency, $\alpha = .89$. The Haptic scale also had a medium level, with interrater $\lambda_6 = .70$, and interitem $\alpha = .83$. The measures were higher for the Visual scale, with interrater $\lambda_6 = .85$, and interitem $\alpha = .89$. In the case of concepts, the Auditory scale had a medium interrater reliability $\lambda_6 = .74$, and interitem $\alpha = .75$; as did the Haptic scale,

with interrater $\lambda_6 = .72$, and interitem $\alpha = .74$; and the Visual scale, too, with interrater $\lambda_6 = .70$, and interitem $\alpha = .72$. Compared to Lynott and Connell (2013), the present reliabilities were lower. However, those norms had over double as much data—namely, seventeen respondents per word, versus eight in the current norms—, and alpha increases with more data. This is relevant for the validity of this replication, as discussed further below.

This survey was presented on an Excel sheet, it was completed online, and it was unpaid. Now the fact that all ratings were valid, and the high reliability of the averages, support the feasibility of non-standard ways of testing. In this case particularly, it probably helped to recruit in person.

3.2. Results

3.2.1. Translation-related results

Starting with the results, we will first analyze the adherence of the translations to the originals in terms of dominant modality. Dominant modality corresponds the highest rated modality for a word. Beginning with the properties, the matching between English and Dutch dominant modalities surpassed 80%. That is, over 80% of the words in the Dutch norms came out with the same modality as the English word on which they were based. Wherever scores for several dominant modalities tied out (it happened with about ten words), only one modality could be kept in order to allow for a number of further analyses.⁴ Therefore, ties were resolved as follows. If any of the tied modalities matched that of the original English word, that modality was maintained. This was possible for all properties. It was done as such because the English score had been a major rationale in the translation process. In contrast, Lynott and Connell resolved all ties at random.

In the case of concepts, the overall adherence of the translations also surpassed 80%. There were about twenty cases of tied dominant modalities, which were resolved as with the properties, except for two cases which were randomly assigned because none of the tied modalities coincided with the English word. The minor 20% divergence in the translation of properties and concepts is likely due, first, to the translation shifts that became necessary in order to render natural-sounding words

⁴ The resolution of tied modalities did not affect the calculations for the adherence of the translations. For any words with tied dominant modalities, all the tied modalities were taken into account.

in Dutch. Secondly, it may be due to the natural semantic asymmetry that holds between similar terms across any languages (i.e., there are no such things as absolute synonyms). Pearson's correlation tests were also performed to check the overlap between both norms (Table 1).

Table 1. Excerpt from the analysis file, abridged. Only items normed in both languages are shown. Key: conc = concept; prop = property; A = Auditory; H = Haptic; V = Visual.

CAT	DUTCH						ENGLISH (Lynott & Connell, 2009, 2013)					
	WORD	MOD	EXC	A	H	V	WORD	MOD	EXC	A	H	V
prop	bolvormig	V	49%	0.60	2.60	4.20	globular	V	43%	0.78	3.17	4.44
prop	bonzend	A	32%	4.00	3.14	1.29	thudding	A	46%	4.57	2.24	2.86
conc	boosheid	V	31%	4.00	1.22	4.11	anger	V	41%	3.71	1.41	4.12
prop	borstelrig	H	34%	1.00	4.00	3.75	bristly	H	37%	1.95	4.65	4.30
conc	bot	H	19%	2.14	3.71	2.57	bone	H	27%	1.59	4.06	3.41
prop	botsend	A	07%	2.50	2.00	2.38	crashing	V	40%	4.57	1.24	4.62
conc	bouw	V	33%	2.67	1.33	4.00	construction	V	39%	2.94	2.06	4.00
conc	bouwer	V	40%	2.56	1.00	4.00	builder	V	38%	3.76	1.24	4.29

Beginning with the properties, auditory strength in English held a large positive correlation with the Dutch one, $r = .795$, $p < .001$. Haptic strength in English and Dutch bore a slightly smaller correlation, $r = .690$, $p < .001$. Visual strength in English and Dutch had a similar correlation, $r = .711$, $p < .001$. The correlation for Exclusivity was smaller, $r = .475$, $p < .001$. In the case of concepts, auditory strength in English held a large positive correlation with Dutch ones, $r = .683$, $p < .001$. Haptic scores in English and Dutch bore a slightly smaller correlation, $r = .624$, $p < .001$. Visual scores in English and Dutch had a similar correlation, $r = .659$, $p < .001$. The correlation for Exclusivity was smaller, $r = .428$, $p < .001$. In all, the broad overlap between English and Dutch norms item by item warrants some item by item comparisons that will be reported—particularly in correlation and ANOVA tests.

3.2.2. Assumption of psychometric invariance

The recent reproducibility crisis in the field of psychology has been sharpening the methods used for replication. For instance, Fabrigar and Wegener (2016) argue for the importance of meta-analytic measures in the validity of (non-)replications. They refer particularly to the variance in the original studies and in their replications, which should hold similar if statistical tests are to be compared. Variance is a building block of statistical tests, as greater variance is penalized for in

the test of significance. This is relevant for both replications and non-replications. If a researcher finds an alleged replication, but the variance in the replication study is greatly smaller or larger than that of the original study, then doubts are cast, because the results of the replication might actually stem from third variables or confounds. Vice versa, a violation of invariance in a non-replication could mean that an otherwise positive replication is only spoiled by a greater variance in the replication. For the present purpose, what the invariance assumption requires is that the two language samples vary similarly across the different levels of their variables, i.e., across the three dominant modalities. Descriptives—particularly means and standard deviations—are the usual indicator for this. The descriptives for the two norms presented in the next section validate this assumption: the figures fluctuate in systematic proportions within the two samples. A good illustration is also found in the Clusterings section, further below.

3.2.3. Descriptives of dominant modality and perceptual strength

The mean scores were entered into the analysis file above described, and coupled with the English scores from Lynott and Connell, as illustrated in Table 1 above.

The overall perceptual strength of the Dutch items is shown in Table 2.⁵

Table 2. Perceptual strength in English and Dutch samples.

	Properties				Concepts			
	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>95% CI</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>95% CI</i>
ENGLISH	3.33	1.35	0.07	0.14	2.72	0.96	0.05	0.10
DUTCH	3.67	0.98	0.05	0.10	3.28	0.96	0.05	0.09

The differences make sense considering the sampling method applied in the different norms. The English and the Dutch norms for concepts are a bit more distant than the property norms because the former were sampled differently. While the concepts in the English norms were compiled regardless of their potential modalities, the Dutch norms were entirely compiled with a view to potential modality, and hence the 0.60 difference in overall perceptual strength.

⁵ For accurate comparisons, the English data were reanalysed without the olfactory and gustatory items.

Turning to the perceptual strength per modality, there was a general dominance of the visual modality. This is reflected in two proportions: the scores on the visual modality were higher overall (Table 3), and there were more dominantly visual words (Table 4).

Table 3. Perceptual strength per modality in English and Dutch samples.

NORMS	Strength	Properties				Concepts			
		<i>M</i>	<i>SD</i>	<i>SE</i>	<i>95% CI</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>95% CI</i>
ENGLISH	Auditory	1.73	1.67	0.09	0.18	2.16	1.09	0.06	0.11
	Haptic	2.41	1.62	0.09	0.17	1.86	1.13	0.06	0.11
	Visual	3.80	1.06	0.06	0.11	3.55	0.80	0.04	0.08
DUTCH	Auditory	1.74	1.29	0.07	0.14	1.97	1.03	0.05	0.11
	Haptic	1.96	1.12	0.06	0.12	1.87	1.13	0.05	0.10
	Visual	3.22	1.15	0.06	0.12	3.13	0.95	0.05	0.09

Table 4. Perceptual strength (0-5) across modalities for each dominant modality, along with modality exclusivity and sample size.⁶ A = Auditory; H = Haptic; V = Visual.

	DOMINANT MODALITY											
	ENGLISH NORMS						DUTCH NORMS					
	Properties			Concepts			Properties			Concepts		
	A	H	V	A	H	V	A	H	V	A	H	V
A strength	4.59	1.12	0.98	3.54	1.35	2.03	3.82	1.37	1.23	3.45	1.52	1.81
H strength	0.70	4.33	2.33	1.03	4.14	1.87	1.22	3.55	1.85	1.50	3.34	1.84
V strength	2.31	3.44	4.41	2.71	3.43	3.67	1.70	2.72	3.75	2.38	2.72	3.30
Exclusivity %	57.4	37.0	48.9	44.1	35.3	39.2	42.8	29.2	41.2	28.1	25.7	29.1
<i>N</i>	68	70	205	42	14	336	64	45	227	48	45	318

This dominance of vision is to a great extent due to the translation, which was done pursuing to keep the modality of the source term. That said, it coincides with other norms (van Dantzig et al., 2011; Winter, Perlman, & Majid, 2018), and with data including conversation across cultures (San Roque et al., 2015), and even sensory perception (Schmid, Büchel, & Rose, 2011). The data from

⁶ We detected a few negligible differences between our own descriptives of the English data and those reported in the original studies. These differences, however, did not extend to more than five measures, and the biggest difference was of 0.04. Our own measures are reported, as they determine the reanalyses.

conversation is very relevant because it indicates that the visual dominance is not necessarily caused by the mode of presentation of the stimuli, which is often visual (cf. Connell & Lynott, 2014).

There are also differences across the two languages. One of the most notable concerns the proportion of haptic concepts relative to the other modalities. While in the Dutch norms this figure is not far from that of auditory concepts (45 versus 48), in the English norms the figure is notably lower (14 versus 42). Another difference regards the proportion of visual items compared to the other modalities, which is greater in the Dutch norms than in the English ones. These differences could not be statistically tested due to the coupling of items across languages (see Supplementary Materials).

3.2.4. Critical results

Modality exclusivity

Modality exclusivity scores are a convenient feature introduced with the Lynott and Connell norms. They are calculated for each word as the range of the three modality scores divided by the sum. This index is possibly comparable to concreteness (though better at predicting behavioural responses; Connell & Lynott, 2012). Thus it may be useful as a corpus measure for psycholinguistic studies in Dutch, for instance on the topic of conceptual processing. Following from the previous table, the distribution of modality exclusivity across categories, modalities and languages are illustrated below by means of stacked bar plots (as in van Dantzig et al., 2011). In these plots, the X axis contains different sub-samples (Auditory concepts, Auditory properties, Haptic concepts, etc). The different colour gradients represent five percentiles of the Exclusivity variable. Finally, the Y axis is based on ‘counts,’ that is, sub-sample sizes (Figure 1).

The two plots look very similar, with the only notable difference of a higher overall exclusivity for English items compared to Dutch items. This was statistically tested. One-sample t-tests—performed on the English data with respect to the Dutch figures—confirmed a significant difference between English ($M = 0.48$, $SD = 0.17$) and Dutch ($M = 0.40$, $SD = 0.18$) properties, $t(342) = 8.70$, $p < .001$, $d_z = 0.47$ (95% $CI = 0.46, 0.50$).

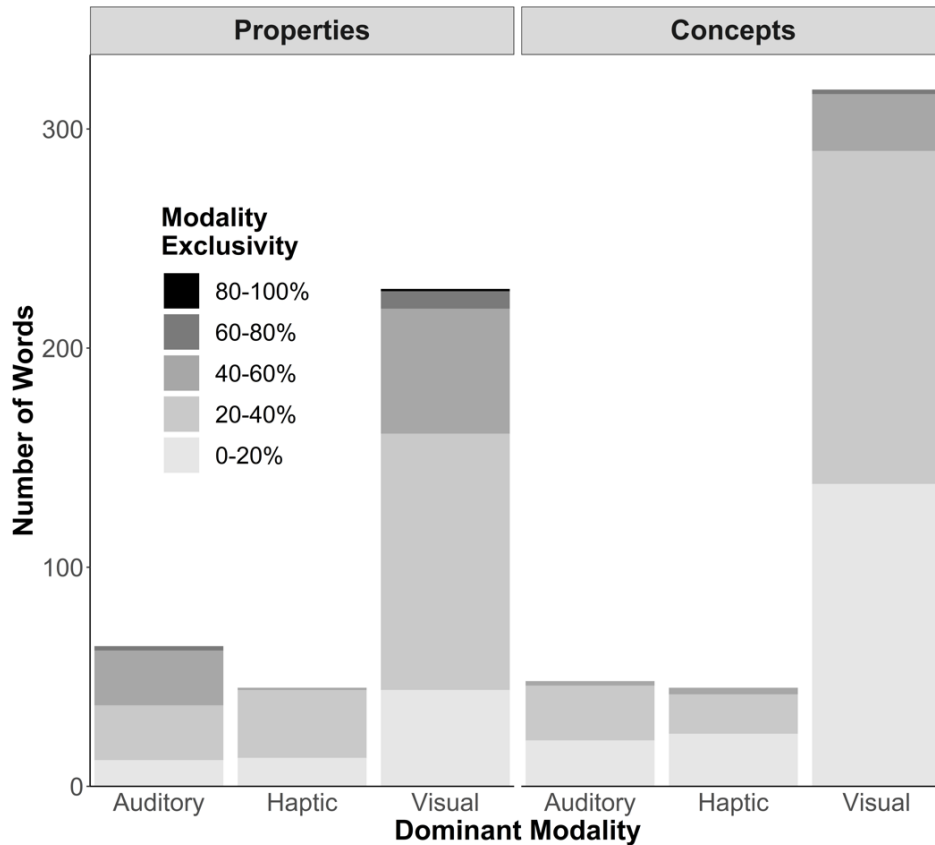


Figure 1. Distribution of modality exclusivity percentiles.

The difference was also significant between English ($M = 0.40$, $SD = 0.12$) and Dutch concepts ($M = 0.29$, $SD = 0.15$), $t(391) = 17.10$, $p < .001$, $d_z = 0.86$ (95% $CI = 0.38, 0.41$). As an effect size, Cohen's d_z was calculated (Lakens, 2013). The greater effect size for the comparison of properties samples across languages suggests that properties are more different across languages than concepts are. This could be partly explained by the sampling of the materials.

Winter (2016) found that the sampling of stimuli may influence their modality exclusivity. Specifically, words selected on the basis of their modality would render greater exclusivity than words selected regardless of any potential modalities. The concepts samples in the two languages were created differently. Whereas the English set was 'sampled,' or randomly created, the Dutch set was created with an eye to potential modalities. In contrast, the properties samples in both norms were both created with an eye to modality. In all, the present data would be in line with Winter's finding.

Sampled versus modality-bound materials

The sampling confound just discussed might have influenced the English norms. While the properties were created with a view to potential modalities, the concepts were sampled. For that reason, the difference in modality exclusivity thereof would have to be regarded with caution. In contrast, the present materials—properties and concepts—were all created *attending* to modality. We will now go on to statistically test the greater exclusivity of properties over concepts, as well as across dominant modalities. Table 5 presents the results of the separate tests on the English and the Dutch norms, including the interaction of the two factors. Eta-squared of the population (η^2_p) is provided as a measure of effect size (Lakens, 2013).

Table 5. ANOVAs on the modality exclusivity across dominant modalities. * $p < .01$; ** $p < .001$.

Source	ENGLISH					DUTCH				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	η^2_p	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	η^2_p
Dominant modality	2	1.26	0.63	31.50**	0.08	2	0.45	0.22	8.67**	0.02
Category	1	1.56	1.56	77.85**	0.10	1	2.42	2.42	93.59**	0.11
Modality:Category	2	0.11	0.05	2.68	0.01	2	0.18	0.09	3.49*	0.01
Residuals	729	14.62	0.02			741	19.12	0.03		

The results are similar for the two languages in general terms. Yet, the dominant modality effect is greater in the English norms. Also, the Dutch norms present an interaction effect of dominant modality and category (concept versus property), which is not corresponded in the English norms.

Planned post-hoc contrasts were performed to check the specific differences across the three dominant modalities. The first contrast was set for auditory versus visual words. The second contrast was for the previous auditory and visual versus haptic words. In the English norms, the first contrast was significant, $F(5, 729) = 29.29, p < .001$. In the Dutch norms, neither contrast was significant, $F(5, 741) = 23.58, p < .001$. There were also important differences for properties and concepts (Table 6). In both English and Dutch, haptic words showed less exclusivity than auditory and visual words within properties, but not within concepts.

In both norms, results confirm the greater multimodality of concepts over properties. Importantly, the fact that a difference arises from properties to concepts within the Dutch norms—where the creation of both sets was modality-bound—underscores this difference. This is despite the fact that

Table 6. Planned contrasts for the previous ANOVAs. * $p < .05$; ** $p < .01$; *** $p < .001$. *Aud* = Auditory; *Hap* = Haptic; *Vis* = Visual; *Cat* = Concept or Property category.

	ENGLISH			DUTCH		
	Estimate	<i>SE</i>	<i>t</i>	Estimate	<i>SE</i>	<i>t</i>
(Intercept)	0.40	0.01	26.73***	0.28	0.02	23.99***
Aud v Vis	0.01	0.01	2.14*	-0.00	0.01	-0.40
Aud-Vis v Hap	-0.02	0.01	-1.60	-0.01	0.01	-1.05
Category	0.08	0.02	4.80***	0.11	0.02	6.38***
Contrast 1 : Cat	0.01	0.01	1.15	0.01	0.01	0.77
Contrast 2 : Cat	-0.03	0.01	-2.22*	-0.03	0.01	-4.64**

there might still be a small influence from the English norms, as the Dutch norms were mainly translated from those (although rated independently).

Overall, the results on modality exclusivity are at one with the nature of human perception. Exclusivity seems to reflect the job of word categories. As Lynott and Connell (2013) pointed out, properties are in charge of creating a (modal) quality, whereas concepts can keep a more passive stance. Exclusivity seems to also reflect the natural distribution of percepts captured by the human senses. Visual and auditory strength would have relatively higher exclusivities because whatever we see or hear often lacks the company of other percepts. That is, we can often see things but not hear or touch them, and by the same token, we often hear things that we cannot see or touch. Now, in contrast, if we can touch something, we likely can see it and hear it too—hence the low exclusivity of haptic items (Connell & Lynott, 2016).

Peer-modalities and learned heuristics

The correlations among modalities and exclusivity from the English norms coincide with those of the Dutch norms, as the correlations below illustrate (Table 7). The visual modality bears a large, positive correlation with the haptic one. In contrast, these two modalities are negatively correlated with the auditory one. The visual and haptic modalities could be regarded as ‘peer-modalities.’ These different relations might be associated to different levels of attention in semantic processing. Louwse and Connell (2011) showed that the peer-modalities—haptic and visual, on the one hand, and gustatory and olfactory, on the other—are peers too in the minds of comprehenders.

Table 7. Correlations among modality strength and exclusivity in the different norms.A = Auditory; H = Haptic; V = Visual. *** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .1$.

		PROPERTIES				CONCEPTS			
		A	H	V	Exc	A	H	V	Exc
ENGLISH	A	–	-.427***	-.625***	.018	–	-.176***	-.008	-.276***
	H		–	.234***	-.621***		–	.554***	-.393***
	V			–	-.053			–	-.065
	Exc				–				–
DUTCH	A	–	-.228***	-.513***	-.173**	–	-.009	.085†	-.410***
	H		–	.193***	-.482***		–	.441***	-.316***
	V			–	.162**			–	.122*
	Exc				–				–

A reanalysis of Lynott and Connell’s modality-switching experiment revealed that shifts across peer-modalities were softer than shifts across non-peer modalities.

Connell and Lynott (2016) contend that linguistic associations such as the so-called peer-modalities could be learnt through linguistic experience, due precisely to the cognitive shortcuts such as that shown by Louwse and Connell. By means of a ‘learned heuristic,’ comprehenders could attend to haptic information even where visual information is the target, or vice versa. Yet, the more common circumstance of those would be to rely on visual information, because—as reported in the above section—visual information is there where haptic information may not be. By the same token, the auditory modality would come in as the least useful due to its relative isolation, where it offers fewer gateways into other modalities. Indeed, this possibility would be supported by the finding that people increasingly sidestep some auditory information as they grow up (Sloutsky & Napolitano, 2003). Such learned heuristics would be especially useful whenever representational capacity is limited, or where the information available is not about the most appropriate modality. Possibly, such learned heuristics would be the basis of language-based semantic processes, such as those in the Symbol Interdependency Theory, and in the Language and Situated Simulation theory (Louwse & Connell, 2011; Santos, Chaigneau, Simmons, & Barsalou, 2011).

Winter (2016) delves into the trade-off that exists between unimodality and multimodality. He notes that the overall multimodality of words contrasts with the unimodal tendency of word combinations. Properties tend to modify concepts from the same modality, or otherwise from a peer-modality. He thus concludes that language has a ‘sweet spot’ between unimodality and multimodality.

Modality as a continuum

In these norms, as in any others based on scaled ratings, some words could be described as unimodal, others as bi-modal, and others as tri-modal or multimodal (Table 8).

Table 8. Examples of unimodal, bimodal, and tri-modal terms. MOD = Dominant Modality; EXC = Modality Exclusivity; A = Auditory; H = Haptic; V = Visual.

DUTCH NORMS							
	DUTCH	ENGLISH	MOD	EXC	A	H	V
Unimodal	gespikkeld	speckled	v	85%	0.25	0.25	4.38
	echoënd	echoing	a	80%	4.63	0.25	0.63
Bimodal	metalen	metal	v	20%	2.00	4.00	4.00
	broos	brittle	h	30%	1.00	3.25	3.25
Tri-modal	knapperig	crisp	h	0%	2.40	2.40	2.40
	pijnvol	aching	h	8%	2.20	2.80	2.80

In spite of these case analyses, however, the labels ‘unimodal,’ ‘bimodal,’ and ‘tri-modal’ lack quantitative warranty for the following reasons. Even though any scale-based norms (i.e., with a rating for each modality) will certainly contain words that are mostly unimodal, and others that are mostly bimodal or tri-modal, the fact remains that words with a modality exclusivity score of 1 or 0 are either absent or extremely rare. Indeed, they occur only twice among the present 747 items, and never in Lynott and Connell (2009, 2013), or in van Dantzig et al. (2011). Modality is clearly a continuum. This leads us to question, how close do different modalities have to be for a word to be called ‘bimodal’ or ‘multimodal?’ We lack some cut-off points. For a quantitative take, readers are directed to the norms file, which may be entirely sorted based on the Exclusivity column.

Modality clusterings

Lynott and Connell (2009, 2013) illustrated the relations among different modalities by means of clusterings. These telling PCA-based plots also enabled a good visualization of the dispersion within each dominant modality. In order to continue comparing the English and the Dutch norms, these plots were reproduced with the English data limited to the relevant modalities, and the Dutch data. For the greatest accuracy, the Dutch PCA included the 24 extra words that were added independently of the English norms, and the English PCA included the 12 words that were not included in the Dutch norms due to redundancy of translations. The property and concept samples were analyzed independently, leading to a total of four analyses. It was done as follows.

For all four plots, preliminary, unrotated analyses with a preset of three factors (the total number of variables) indicated that two components should be extracted, on the basis of Joliffe's threshold—i.e., eigenvalue $> .7$. This was confirmed by scree plots (Field, Miles, & Field, 2012). The same had come up in Lynott and Connell (2013). Next, the definitive analysis was performed through a varimax (orthogonally) rotated PCA with Kaiser normalization, where two components were preset. In the English properties analysis, the extracted factors commonly explained 89% of the variance, while the factors in the concepts analysis explained 86%. For the Dutch properties, the extracted factors explained 84% of the variance, while the factors in the concepts analysis explained 82%. Table 9 further shows the correspondence of factors to original variables in the four analyses. Scores indicating adherence to a component are marked in bold. These correspondences are not just based on a naked eye observation: any correlation coefficients above $.7$ indicate that 'at least half of the variance in a variable is explained by the component' (Lynott & Connell, 2013, p. 523).

Table 9. Correlations between variables and components (no significance tested). Correlations above $r = .7$ (i.e., 50% variance explained) in bold.

	ENGLISH NORMS				DUTCH NORMS			
	Properties		Concepts		Properties		Concepts	
	RC1	RC2	RC1	RC2	RC1	RC2	RC1	RC2
Auditory	-.825	-.360	-.040	.990	-.852	-.158	.030	.994
Haptic	.156	.977	.865	-.201	.107	.993	.854	-.090
Visual	.932	.040	.894	.090	.872	.080	.844	.120

These coefficients underscore the similarity between the English and the Dutch samples, with the adherence of variables to components matched in the two languages. Particularly, the properties samples present a strong opposition of the auditory modality against the other two, whereas the concepts samples present a lesser opposition, allowing separate components for the auditory modality, on the one hand, and the visual and haptic modalities, on the other.

The two rotated factors extracted were then plotted on X and Y. As the clusterings in Figure 2 illustrate, properties in both languages are perceptually stronger than concepts. These plots also underscore the assumption of psychometric invariance referred to above. Variance fluctuates similarly in the original study and in the replication—i.e., across the three modalities, as well as for properties versus concepts. Furthermore, the plots underscore the extent of the overlap between the English and the Dutch norms, as the three modalities are located in very similar areas.

Sound symbolism

Sound symbolism is an interesting psycholinguistic phenomenon: words often sound like what they mean. Put another way, the sound of some words bears a non-arbitrary relation to their meaning. In a broader form, the effect is also known as iconicity (Dingemans, Blasi, Lupyan, Christiansen, & Monaghan, 2015). Lynott and Connell (2013) analyzed whether the words in their norms reflected sound symbolism. They hypothesized that, if sound was indeed an integral part of the meaning of words, then auditory experience should be the best predictor of the lexical features of those words. Further, because auditory scores are barely or negatively related to haptic and visual ones, any effects of Auditory strength should pull in the opposite direction from the other modalities. Interestingly, sensory experience has been shown to be the best predictor of iconicity, over imageability, frequency and systematicity (Winter, Perlman, Perry, & Lupyan, 2017).

The same hypotheses posited by Lynott and Connell would hold for the Dutch norms (see again Table 7 for the correlations among modalities). For the testing, the perceptual strength of all three modalities served together as independent variables, which predicted measures of word length, distinctiveness and frequency, the latter separate. The sources for those lexical measures are listed next. Since some words had missing cases on some variables, the percentage of items with measures is specified after each variable.

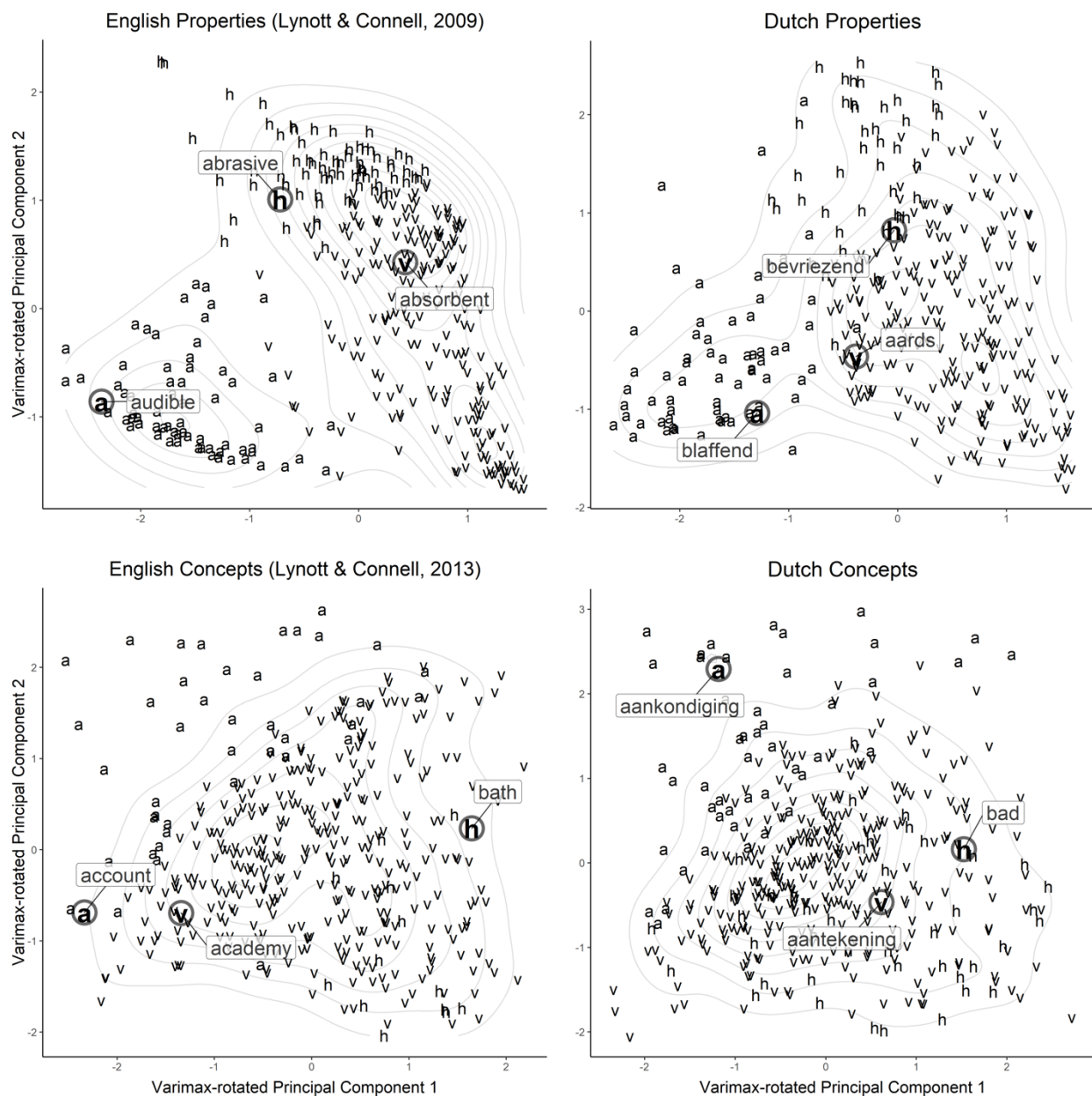


Figure 2. PCA-based clusterings for English and Dutch properties and concepts. Letters show the dominant modality of each word, while the contours reflect the dispersion within each modality.

First, the online database CLEARPOND (Marian, Bartolotti, Chabal, & Shook, 2012) was used to estimate number of phonemes (available for 55% of properties and 95% of concepts), as well as phonological and orthographical neighbourhood sizes (all properties, all concepts). Two similar frequency measures—log-10 word frequency and log-10 corpus diversity—were retrieved from SUBTLEX-NL (available for 86% of properties and 99% of concepts; corpus from Keuleers,

Brysbaert, & New, 2010). The CELEX lemma frequency per million was also retrieved (available for 74% of properties and 97% of concepts; corpus from Baayen, Piepenbrock, & van Rijn, 1993). In addition to those, measures for age of acquisition and concreteness were retrieved (available for 69% of properties and 97% of concepts; corpus from Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014). In regard to the adjectives, all the above measures are for the neuter form (i.e., for words taking the article ‘het’). The norms also contain the log-10 SUBTLEX-NL corpus diversity of inflected adjectives.

Normality of distribution was tested for all variables, and all turned out to be skewed or kurtosed. Solutions were sought via three different transformations: log, square, and square root. None of those improved the distributions far enough, so no transformation was applied. Next, correlation tests were conducted with all lexical variables in the properties and in the concepts samples (Tables 10 and 11).

Table 10. Intercorrelations of lexical variables in the Dutch properties sample.

Phon. = Phonological; *Orth.* = Orthographic; *Neigh.* = Neighbourhood Size; *AoA* = Age of Acquisition. ** $p < .001$; * $p < .05$.

DUTCH PROPERTIES											
	Letters	Phonemes	Phon. Neigh.	Orth. Neigh.	Context. Diversity	Word freq.	Lemma freq.	Age of acquis.	Concrete- ness	<i>M</i>	<i>SD</i>
Letters	–	.940**	-.727**	-.703**	-.508**	-.509**	-.550**	.405**	-.090	7.12	2.26
Phonemes		–	-.716**	-.732**	-.486**	-.486**	-.555**	.457**	-.090	5.38	1.95
Phonolog. Neighbou.			–	.895**	.467**	.470**	.518**	-.417**	.118	4.57	7.56
Orthogra. Neighbou.				–	.477**	.478**	.517**	-.438**	.156*	3.32	4.87
Context. diversity					–	.995**	.838**	-.654**	-.166*	1.80	1.02
Word frequency						–	.832**	-.646**	-.161*	1.88	1.09
Lemma frequency							–	-.700**	-.100	1.08	0.80
AoA								–	-.254**	7.97	2.13
Concrete.									–	3.27	0.70

Table 11. Intercorrelations of lexical variables in the Dutch concepts sample.

Phon. = Phonological; *Orth.* = Orthographic; *Neigh.* = Neighbourhood Size. ** $p < .001$; * $p < .05$.

DUTCH PROPERTIES											
	Letters	Phonemes	Phon. Neigh.	Orth. Neigh.	Context. Diversity	Word freq.	Lemma freq.	Age of acquis.	Concrete- ness	<i>M</i>	<i>SD</i>
Letters	–	.942**	-.647**	-.630**	-.364**	-.381**	-.212**	.491**	-.415**	6.71	2.54
Phonemes		–	-.617**	-.633**	-.362**	-.369**	-.237**	.513**	-.397**	5.82	2.18
Phonolog. Neighbou.			–	.879**	.329**	.338**	.201**	-.467**	.391**	5.53	8.27
Orthogra. Neighbou.				–	.349**	.352**	.220**	-.437**	.348**	4.03	5.73
Context. diversity					–	.987**	.776**	-.585**	.007	2.66	0.66
Word frequency						–	.757**	-.601**	.039	2.85	0.76
Lemma frequency							–	-.430**	-.124**	1.54	0.64
AoA								–	-.569**	8.07	1.07
Concrete.									–	3.02	1.07

The correlations in both samples confirm the adherence of the variables within each of the three groups. In contrast to those, the extra variables Age of Acquisition and Concreteness had much smaller correlations, with only one coefficient reaching the key threshold of .7, which indicates half of the variance explained. They were still maintained for the sound symbolism analysis, but were spared from the Principal Components Analysis reported below.

Principal Components Analysis. Lynott and Connell (2013) found that the three groups of lexical variables—length, distinctiveness and frequency—were intercorrelated. Specifically, long words were related to lower frequencies and greater distinctiveness in sound and spelling (i.e., smaller neighbourhood size). This intercorrelation would entail that, even if these variable are entered into different regressions they would still be affected by each at core. In order to isolate each group, the authors turned to Principal Components Analysis. They entered all of the variables from the three groups together into this analysis, and let the system arrange whatsoever groups. The result was clear: there were three components, and their contents corresponded exactly to the groups of

lexical variables entered. This analysis was reproduced hereby with the Dutch properties and concepts separately.

The PCA yielded three components, which corresponded exactly to each of the groups of variables entered, namely, those for length, those for frequency, and those for distinctiveness. It was done in a similar way to the PCAs reported above. Separate PCAs for properties and concepts were conducted. At a first stage, a preliminary analysis was run to check how many factors should be selected in the definitive analysis. Whereas Lynott and Connell ran this probe with unrestricted factors, we ran it with seven factors, i.e., the total number of variables. This was preferred because it would allow us to look at eigenvalues as an indicator, besides the scree plot. For properties as well as concepts, Joliffe's threshold—i.e., eigenvalue $> .7$ —indicated that three components should be extracted. Scree plots again underscored that three components should be extracted. This was further confirmed by explained variance in further restricted PCAs, as any more components than three would only explain a negligible amount of variance. The definitive analysis then was based on a varimax (orthogonally) rotated PCA with Kaiser normalization, where three components were preset. Table 12 presents the resulting correlations, which show that the components correspond to the different groups of lexical variables.

Table 12. Correlations between variables and components (no significance tested). *Phon.* = Phonological; *Orth.* = Orthographic. Correlations above $r = .7$ (i.e., 50% variance explained) in bold.

Lexical variable	PROPERTIES			CONCEPTS		
	RC1	RC2	RC3	RC1	RC2	RC3
Letters	.862	-.350	-.332	.910	-.185	-.334
Phonemes	.873	-.246	-.392	.917	-.163	-.326
Phon. neighbourh.	-.368	.294	.858	-.322	.169	.903
Orth. neighbourh.	-.355	.309	.859	-.317	.148	.901
Context. diversity	-.252	.928	.227	-.181	.945	.161
Word frequency	-.250	.927	.233	-.192	.937	.168
Lemma frequency	-.244	.852	.309	-.050	.894	.070

Regressions. Next, the regressions were conducted. As in Lynott and Connell (2013), separate regressions (twelve in this case) were run with each lexical variable and each rotated component as a dependent variable, which were predicted by the three modality scores. Both the dependent

and the independent variables were standardized (mean-centered and scaled) prior to the regression (see the analysis on English concepts in Lynott & Connell, 2013, Table 7 above). This was done particularly to facilitate the comparison with the English norms, which were standardized too. None of the regressions had the problem of multicollinearity, with the largest VIF < 10, mean VIF \approx 1, and tolerance > .2 (Field, Miles, & Field, 2012). Results are presented in Table 13, and graphically illustrated in Figures 3 and 4. The portion with the largest data basis is that of the Rotated Components. Again on this aspect, the Dutch norms reproduced the English ones. Auditory strength set itself apart from the other modality scores in predicting the lexical variables.

Table 13. Separate stepwise regressions with each lexical variable as DV and the three perceptual strengths as IVs (Dutch norms). *Phon.* = Phonological; *Orth.* = Orthographic; *Neigh.* = Neighbourhood Size; *AoA* = Age of Acquisition. Standardized (β) coefficients. Bidirectional selection, with automatic inclusion and exclusion. $N_{\text{properties}} = 336$. $N_{\text{concepts}} = 411$ (save missing lexicals). *** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .1$.

Variable	PROPERTIES			CONCEPTS		
	A	H	V	A	H	V
Length						
Letters	+0.11*	-0.09	(-0.02)	+0.05**	-0.07***	(-0.02)
Phonemes	+0.26*	(-0.09)	(+0.11)	+0.06***	-0.06**	(-0.02)
Distinctiveness						
Phon. neigh.	-0.06**	(+0.02)	(-0.01)	-0.04*	+0.11***	(+0.01)
Orth. neigh.	-0.05*	(+0.01)	(-0.01)	-0.05*	+0.11***	(+0.02)
Frequency						
Context. diversity	-0.07**	(-0.00)	+0.07**	+0.20***	(+0.06)	+0.11**
Word frequency	-0.07**	(-0.00)	+0.07**	+0.20***	(+0.05)	+0.12**
Lemma frequency	(+0.05)	(-0.03)	+0.15†	+0.15**	(-0.05)	+0.12*
RCs						
Length	+0.06†	(-0.02)	(+0.03)	+0.05***	-0.04*	(-0.01)
Distinctiveness	-0.05	(+0.01)	-0.08**	-0.03*	+0.07***	(-0.01)
Frequency	+0.29*	(-0.07)	+0.19*	+0.23***	(-0.04)	+0.10†
Other variables						
Concreteness	-0.17†	(+0.07)	+0.18*	-0.05***	+0.11***	+0.13***
AoA	(+0.04)	(-0.01)	-0.21**	-0.10*	-0.26***	-0.29***

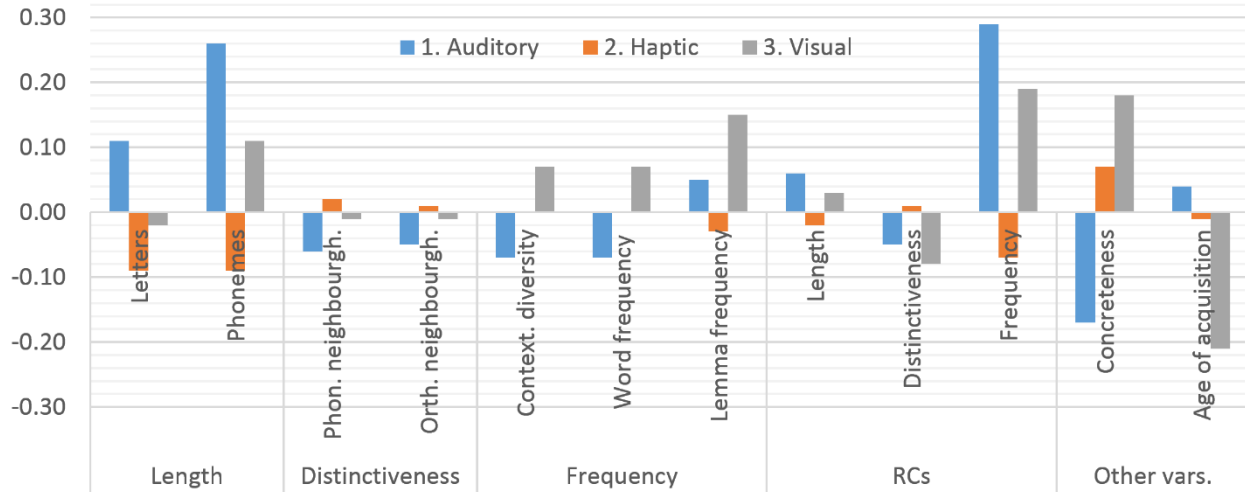


Figure 3. Graphical representation of the sound symbolism regressions for Dutch properties.

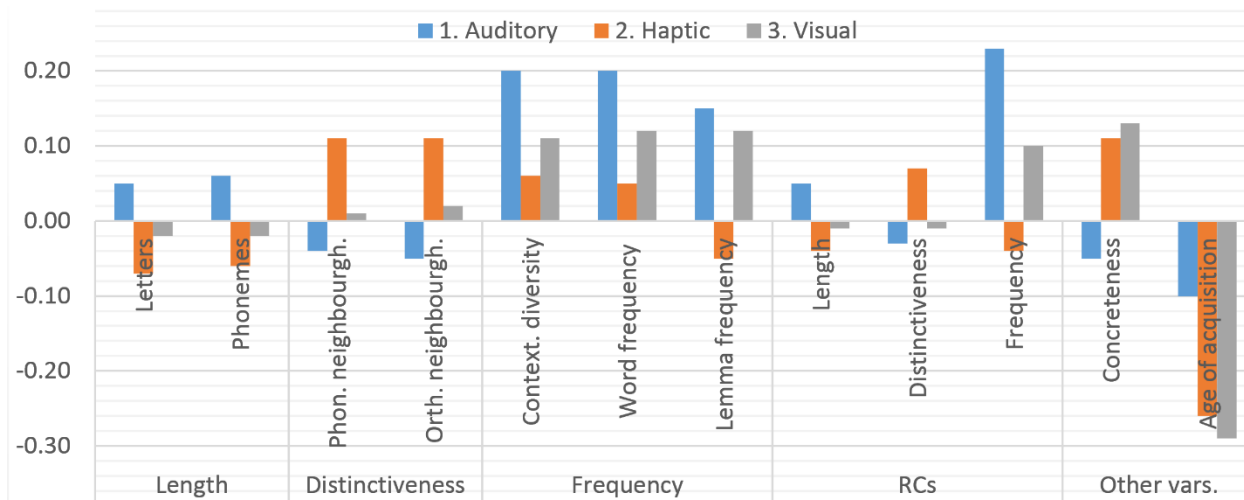


Figure 4. Graphical representation of the sound symbolism regressions for Dutch concepts.

Specifically, auditory scores tended to either bear more power than the other two modalities, or else to pull in the opposite direction from the strongest modality. The direction refers to the polarity of the regression coefficient. In the properties sample, the uniqueness of Auditory scores only tapered off when predicting the rotated component for Distinctiveness. In the concepts sample, the exception occurred in the prediction of age of acquisition. This particular case might not be coincidental, as a negative relationship between iconicity and age of acquisition has appeared in both English and Spanish (Perry, Perlman, & Lupyan, 2015).

The greater predictive power of auditory scores is in line with an interesting finding on the topic. Dingemanse et al. (2016) showed that words with iconic associations in the auditory realm are easier to guess than words with associations in other modalities. Since the iconic association is based on the sound of the word, a switching cost will arise when a different modality is triggered.

This analysis of sound symbolism contrasts with more controlled measures such as those based on particular phonetic properties, or others based on languages that are better known for their sound-symbolism (see Lockwood & Dingemanse, 2015). To be sure, the concepts analyzed in Lynott and Connell (2013), and the concepts and properties in the current study, are only controlled at a superficial level, so this is a big-data kind of analysis. Thus, questions such as the relative sound-symbolism across lexical categories or dominant modalities could not be ascertained (cf. Perry, Perlman, & Lupyan, 2015; Winter, Perlman, Perry, & Lupyan, 2017). Nevertheless, a trend is clearly visible within Dutch properties and concepts, which furthermore converges with English words (Lynott & Connell, 2013). Further research could re-address this method in order to confirm that the regression applied is indeed indicative of sound symbolism, and not of any alternative phenomena. Also, if the present findings are on the right track, we should further explore sound symbolism from the standpoint of modality, at best focused on languages in which sound symbolism has received less attention.

4. General discussion

Research on the different aspects of cognition is at the mercy of experimental stimuli. These materials usually have to be as controlled as possible, which includes their size, position, frequency time of presentation, etcetera, down to the greatest detail. Such controls are often matched by advanced technology—for instance, high-refresh monitors. In the case of language research, that bare minimum of control is often extended by a particular requirement. Each and every item has to be validated by the authority of speakers. This ‘norming’ of linguistic terms will often make a study in itself, or indeed several of them.

For example, in the area of conceptual modality, subsequent studies have competed for explaining behavioural results better and better. This is not gratuitous, but for some experimental designs in language science, such a dedication to the stimulus becomes indispensable. And this trend of control only keeps increasing (Levelt, 2014; Grondelaers, Geeraerts & Speelman, 2007). In a way,

it is the nature of traditional experimental designs: once you implement stimuli that do not necessarily resemble the real world, the more constraints the better. In stark contrast to this, yet, modern approaches have demonstrated the feasibility of using real-life language in experiments (Hartung, Burke, Hagoort, & Willems, 2016; Willems, 2015).

The stimuli, methods and analyses in these modality exclusivity norms are partly based on previous English norms (Lynott & Connell, 2009; Lynott & Connell, 2013). Particularly, the Dutch stimuli were translated from the English ones. There were important design differences: Instead of the five modalities tested in the English study, we narrowed down to the auditory, haptic, and visual modalities. Also, the present norms are for both properties and concepts. The creation of both sets was modality-bound, with all translations attending to the modality of the source term. We re-tested several trends previously found for English. For this comparison, the English data was reanalyzed where necessary, and constrained to the three relevant modalities. This yielded a robust reproduction of their findings. First, visual-dominant words were by far the most numerous, converging with previous modality norms (Lynott & Connell, 2009; 2013; Winter, Perlman, & Majid, 2018), and other data including conversation across cultures (San Roque et al., 2015), and even sensory perception (Schmid, Büchel, & Rose, 2011). This point, however, is determined by the translation process, which attended to the meaning and the dominant modality of the source terms.

More interestingly, visual and haptic perceptual strength were quite related, whereas the auditory one came out as more isolated. These different levels of exclusivity in the language may possibly be associated to different levels of detail in semantic processing (Louwerse & Hutchinson, 2012; Louwerse & Connell, 2011; Simmons, Hamann, Harenski, Hu, & Barsalou, 2008). Functionality, timing, and cortical brain distribution have all been tackled, yet still further research seems necessary to fully understand the cognitive implementation of language statistics alongside perceptual simulation. Third, the three modalities also presented differences in modality exclusivity, with auditory and visual words showing greater unimodality, and haptic words showing greater multimodality. The explanation for this concerns the human perceptual senses: touch is the less powerful sense out of the three, as it doesn't allow us to feel at the distance. Yet, when we can touch something, we can often see and hear it too. Fourth, properties were more unimodal than concepts. Fifth, the data presented sound-symbolism, that is, non-arbitrary

relations between meaning and sound. As such, auditory experience predicted lexical properties of the words better than the other two modalities, or else with a different polarity. This held for both Dutch properties and concepts. Further exploration of this phenomenon through the lens of modality is encouraged.

Acknowledgements

This project would not have been possible without the supervision of Max Louwerse and Roel Willems, the generous help of Wendy Leijten with the translations from English into Dutch, or the contribution from the respondents. Many thanks also to Bodo Winter and Diane Pecher for valuable feedback.

Supplementary materials

All the data and analysis code are available at <https://osf.io/brkjw/>.

References

- Ambrosi, S., Kalenine, S., Blaye, A., & Bonthoux, F. (2011). Modality switching cost during property verification by 7 years of age. *International Journal of Behavioral Development, 35*(1), 78–83.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX Lexical Database* [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Baker, M. (July 20, 2016). Dutch agency launches first grants programme dedicated to replication. *Nature News*. Retrieved from <http://www.nature.com/news/dutch-agency-launches-first-grants-programme-dedicated-to-replication-1.20287>.
- Barrós-Loscertales, A., González, J., Pulvermüller, F., Ventura-Campos, N., Bustamante, J. C., Costumero, V., Parcet, M. A., & Ávila, C. (2012). Reading salt activates gustatory brain regions: fMRI evidence for semantic grounding in a novel sensory modality. *Cerebral Cortex, 22*(11), 2554–2563.
- Bernabeu, P., Willems, R. M., & Louwrese, M. M. (2017). Modality switch effects emerge early and increase throughout conceptual processing: evidence from ERPs. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 1629-1634). Austin, TX: Cognitive Science Society. <https://mindmodeling.org/cogsci2017/papers/0318/>
- Binder, J. R., Conant, L. C., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology, 33*(3-4), 130–174.
- Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica, 150*, 80-84.
- Burenhult, N., & Majid, A. (2011). Olfaction in Aslian ideology and language. *The Senses & Society, 6*, 19-29.
- Collins, J., Pecher, D., Zeelenberg, R., & Coulson, S. (2011). Modality switching in a property verification task: An ERP study of what happens when candles flicker after high heels click. *Frontiers in Cognition, 2*, 10.
- Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition, 125*, 452-465.
- Connell, L., & Lynott, D. (2014). I see/hear what you mean: Semantic activation in visual word recognition depends on perceptual attention. *Journal of Experimental Psychology: General, 143*, 527-533.

- Connell, L., & Lynott, D. (2016). Do we know what we're simulating? Information loss on transferring unconscious perceptual simulation to conscious imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1218-1232.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 3, 297-334.
- Dingemanse, M., Blasi, D. E., Lupyán, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity and systematicity in language. *Trends in Cognitive Sciences*, *19*, 10, 603-615.
- Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*, *92*, 2, e117-e133.
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429-492.
- Fabrigar, L. R., & Wegener, D. T. (2016). Conceptualizing and evaluating the replication of research results. *Journal of Experimental Social Psychology*, *66*, 68-80.
- Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: Sage.
- Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on 'Estimating the reproducibility of psychological science'. *Science*, *351*, 1037.
- Glenberg, A. M. (1984). A retrieval account of the long-term modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 16-31.
- González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage*, *32*, 906-912.
- Grondelaers, S., Geeraerts, D., & Speelman, D. (2007). A case for a cognitive corpus linguistics. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, & M. J. Spivey (Eds.), *Methods in Cognitive Linguistics*, pp. 149-69. Amsterdam: John Benjamins.
- Hald, L. A., Marshall, J.-A., Janssen, D. P., & Garnham, A. (2011). Switching Modalities in A Sentence Verification Task: ERP Evidence for Embodied Language Processing. *Frontiers in Psychology*, *2*.
- Hartung, F., Burke, M., Hagoort, P., & Willems, R. M. (2016). Taking perspective: Personal pronouns affect experiential aspects of literary reading. *PLoS One*, *11*(5), e0154732.
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*, 643-650.

- Kline, P. (1999). *The handbook of psychological testing* (2nd ed.). London: Routledge.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863.
- Levelt, W. J. M. (2013). *A history of psycholinguistics: The pre-Chomskyan era*. Oxford: Oxford University Press.
- Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: a review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology, 6*, 1246.
- Louwerse, M., & Connell, L. (2011). A taste of words: linguistic context and perceptual simulation predict the modality of words. *Cognitive Science, 35*, 2, 381-98.
- Louwerse, M., & Hutchinson, S. (2012). Neurological evidence linguistic processes precede perceptual simulation in conceptual processing. *Frontiers in Psychology, 3*.
- Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods, 41*, 2, 558-564.
- Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior Research Methods, 45*, 516-526.
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE, 7*, 8: e43230.
- Marques, J. M. (2006). Specialization and semantic organization: Evidence for multiple semantics linked to sensory modalities. *Memory & Cognition, 34*, 60-67.
- Newman, S. D., Klatzky, R. L., Lederman, S. J., & Just, M. A. (2005). Imagining material versus geometric properties of objects: an fMRI study. *Cognition and Brain Research, 23*, 235-246.
- Nuyts, J. (2001). *Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective*. Amsterdam: Benjamins.
- Ondobaka, S., Hald, L., & Bekkering, H. (2016). Embodied predictive processing in social understanding. In M. H. Fischer & Y. Coello (Eds.), *Conceptual and Interactive Embodiment*, pp. 200-216. London: Routledge.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716.
- Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science, 14*, 119-124.

- Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic Bulletin & Review*, *11*, 164-167.
- Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and Its Relation to Lexical Category and Age of Acquisition. *PLoS ONE*, *10*, 9: e0137147.
- San Roque, L., Kendrick, K. H., Norcliffe, E., Brown, P., Defina, R., Dingemanse, M., Dirksmeyer, T., Enfield, N. J., Floyd, S., Hammond, J., Rossi, G., Tufvesson, S., van Putten, S., & Majid, A. (2015). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics*, *26*, 31-60.
- Santos, A., Chaigneau, S. E., Simmons, W. K., & Barsalou, L. W. (2011). Property generation reflects word association and situated simulation. *Language and Cognition*, *3*, 1, 83-119.
- Scerrati, E., Lugli, L., Nicoletti, R., & Borghi, A. M. (2016). The Multilevel Modality-Switch Effect: What Happens When We See the Bees Buzzing and Hear the Diamonds Glistening. *Psychonomic Bulletin & Review*, *24*, 798–803.
- Schmid, C., Büchel, C., & Rose, M. (2011). The neural basis of visual dominance in the context of audio-visual object processing. *NeuroImage*, *55*, 1, 304-311.
- Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology*, *102*, 1, 106.
- Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*, 2802-2810.
- Sloutsky, V. M., & Napolitano, A. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development*, *74*, 822-833.
- Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, *32*, 244-259.
- Speed, L. J., & Majid, A. (2017). Dutch modality exclusivity norms: Simulating perceptual modality in space. *Behavior Research Methods*, *49*(6), 2201–2218.
- Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, *63*, 330-336.
- Turatto, M., Benso, F., Galfano, G., & Umiltà, C. (2002). Non-spatial attentional shifts between audition and vision. *Journal of Experimental Psychology: Human Perception & Performance*, *28*, 628-639.

- van Dantzig, S., Cowell, R. A., Zeelenberg, R., & Pecher, D. (2011). A sharp image or a sharp knife: Norms for the modality exclusivity of 774 concept-property items. *Behavior Research Methods*, *43*, 145-154.
- van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2008). Perceptual Processing Affects Conceptual Processing. *Cognitive Science*, *32*, 579-590.
- van den Bosch, A. (2008). *Het volgende woord*. Inaugural address at Tilburg University. Retrieved from <http://ilk.uvt.nl/hetvolgendewoord.html>.
- Willems, R. M. (2015). *Cognitive neuroscience of natural language use*. Cambridge, United Kingdom: Cambridge University Press.
- Winter, B. (2016). *The Sensory Structure of the English Lexicon*. PhD thesis, University of California, Merced.
- Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, *179*, 213–220.
- Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic? Iconicity in English sensory words. *Interaction Studies*, *18*(3), 443–464.
- Woodruff, D., & Wu, Y. (2012). Statistical considerations in choosing a test reliability coefficient. *ACT Research Report Series*, *10*.