Pablo Bernabeu

Modality exclusivity norms for 747 properties and concepts in Dutch: a replication of English

This study is a cross-linguistic, conceptual replication of Lynott and Connell's (2009, 2013) modality exclusivity norms. Their English properties and concepts were translated into Dutch, then independently tested as follows. Forty-two respondents rated the auditory, haptic, and visual strength of those words. Mean scores were then computed, with a high interrater reliability and interitem consistency. Based on the three modalities, each word also features a specific modality exclusivity, and a dominant modality. The norms also include external measures of word frequency, length, distinctiveness, age of acquisition, and known percentage. Starting with the results, unimodal, bimodal, and tri-modal words appear. Visual and haptic experience are quite related, leaving a more independent auditory experience. These different relations are important because they may correlate with different levels of detail in word comprehension (Louwerse & Connell, 2011). Auditory and visual words tend towards unimodality, whereas haptic words tend towards multimodality. Likewise, properties are more unimodal than concepts. Last, the 'sound symbolism' hypothesis was tested by means of a regression: Auditory strength predicts lexical properties of the words (frequency, distinctiveness...) better than the other modalities do, or else with a different polarity.

Number of words in the manuscript: 11,109. Appendices from page 34

Some words such as properties and concepts bear sensory associations, which can actually be measured. For example, if you go and 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 and Barsalou, 2003; also see Pecher, Zeelenberg, & Barsalou, 2004; Solomon & Barsalou, 2004; Newman, Klatzky, Lederman, & Just, 2005; Marques, 2006; Gonzalez et al., 2006; Vermeulen, Niedenthal, & Luminet, 2007; Lynott & Connell, 2009; Ambrosi, Kalenine, Blaye, & Bonthoux, 2011; Collins, Pecher, Zeelenberg, & Barsalou, 2011; Hald, Marshall, Janssen, & Garnham, 2011).

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. 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.

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; Winter, 2016). 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. Whichever be sufficient will depend on the research question.

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 what extent they experienced each word through the senses of hearing, touch, and vision, on scales from 1 to 5. 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÷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.

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, 2016). Second, visual and haptic experience were quite related, leaving auditory experience as more independent (see also Winter, submitted). These different relations are important because they may correlate with different levels of detail in word comprehension (Louwerse and Connell, 2011). Third, 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, submitted). 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.

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. 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; Sutton & Majid, 2016). Just as we cannot assume the consistency of behaviour across or even within cultures, we should be wary too 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, 2015; see also Baker, 2016).[2] 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 Louwerse 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

[1] Materials retrieved from http://www.lancaster.ac.uk/people/connelll/lab/norms.html.

[2] For an interview about this 'turmoil' and the current resolutions, go to: http://www.sciencefriday.com/segments/the-replication-game-how-well-do-psychology-studies-hold-up/

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.

Method

Respondents

The modality rating was completed by forty-two university students of Radboud University and Tilburg University, Netherlands, who were not paid.[3]

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.[4] On top of that, twenty-four concepts were added, each of which was created to potentially adhere to a particular modality.[5] 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. This is signaled in the analysis file (see Appendices 1 and 2).

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

---

[3] The recruitment developed as follows. At a first stage, about sixty undergrads of the Tilburg School of Humanities, native of speakers of Dutch, were told about the possibility to complete a forty-minute survey for a master's student. About twelve students volunteered, and eventually six did return the survey completed. At a second stage, about two hundred students were approached in person on the Tilburg and Radboud campuses, mainly at the libraries. Every student received an information sheet with the nature of the request—a twenty-minute language survey for a master's student—, and an email address to contact in case of interest. With this sheet, most students also received some candy. During the next days, about 130 of the students failed to establish contact. About seventy students did make contact, and a final forty-two students sent back the survey completed.

[4] Besides the olfactory and gustatory modalities taken out, twelve words from the relevant modalities (seven properties, five concepts) were not included 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.

[5] Done in order to fulfill the stimuli of another experiment in preparation by the current authors.

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,’ 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 Appendix 2).

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.[6] Respondents completed the survey at home or wherever they chose, over the course of a fortnight.

Design and analysis

In the analysis file, Dutch and English data was described in separate columns mostly (see Appendices 1 and 2). 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).

[6] 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.’

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 nine. 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 all analysis code with results in Appendix 3). Two measures were calculated, both based on Cronbach's alpha (Cronbach, 1951). While 'interitem consistency' measures the fit among items independently of raters, 'interrater reliability' does the opposite, measuring the fit among raters, independently of items. 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 G6—was fine too, with all scales above $\alpha = .70$, and an average $\alpha = .79$. The individual figures were as follows. In the case of properties, the Auditory scale had a medium interrater reliability, $\alpha = .78$, and interitem consistency, $\alpha = .89$. The Haptic scale also had a medium level, with interrater $\alpha = .70$, and interitem $\alpha = .83$. The measures were higher for the Visual scale, with interrater $\alpha = .85$, and interitem $\alpha = 89$. In the case of concepts, the Auditory scale had a medium interrater reliability $\alpha = .74$, and interitem $\alpha = 75$; as did the Haptic scale, with interrater $\alpha = .72$, and interitem $\alpha = .74$; and the Visual scale, too, with interrater $\alpha = .70$, 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.

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. To start 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. Further, beyond word-to-word relations, the proportion of items across modalities is also fairly similar in the Dutch and English norms (see the 'N' row in Table 3). 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.[7] 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 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).

Correlation tests (Pearson's for all henceforth) were also performed to check the overlap between both norms (see descriptives in Table 1). To start with the properties, auditory strength in English

[7] 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.

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.

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 spoilt 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 too is in the Clusterings section, further below.

Descriptives for 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 (henceforth, A = Auditory; H = Haptic; V = Visual; conc = concept; prop = property) (see Appendices 1 and 2).

Table 1 Excerpt from the analysis file, abridged. Only items normed in both languages are shown.

	DUTCH						ENGLISH (Lynott & Connell, 2009, 2013)					
CAT	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	borstelig	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

There was a general dominance of the visual modality. This is reflected in two proportions, two sides of the same coin: the scores on the visual modality are higher overall (Table 2), and there are more dominantly visual words (Table 3).

Table 2 Descriptives for English and Dutch samples.[8]

		Properties				**Concepts**			
NORMS	**Strength**	*M*	*SD*	*SE*	*95% CI*	*M*	*SD*	*SE*	*95% CI*
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

This dominance of vision is to a great extent due to the translation, which pursued precisely keeping the same modality of the source term. That said, it coincides with other norms (van

[8] The English data are reanalysed without the olfactory and gustatory items, hence the differences from the figures in Lynott and Connell (2009, 2013)

Table 3 Perceptual strength (0-5) across modalities for each dominant modality, along with modality exclusivity and sample size.[9]

	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
N	68	70	205	42	14	336	64	45	227	48	45	318

Dantzig et al., 2011; Winter & Perlman, 2016), and with data including conversation across cultures (San Roque et al., 2015), and even sensory perception (Schmid, Büchel, & Rose, 2011). The data from 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 and 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 Appendix 2).

Critical results

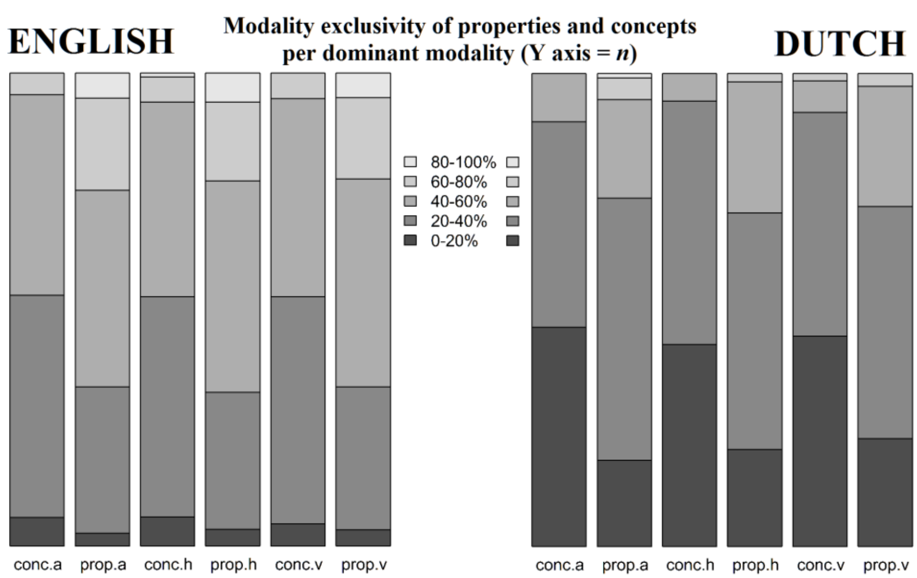
Modality exclusivity

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

[9] A few negligible differences were detected 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.

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).

Figure 1 Distribution of modality exclusivity percentiles



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). 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 (submitted) 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 4 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 4 ANOVAs on the modality exclusivity across dominant modalities.

	ENGLISH					DUTCH				
Source	*df*	*SS*	*MS*	*F*	η^2_p	*df*	*SS*	*MS*	*F*	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		

*$p < .01$; **$p < .001$

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 used 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 two groups versus haptic words. For the English norms, the first contrast is significant, $F(5, 729) = 29.29$, $p <.001$. For the Dutch norms, neither contrast is significant, $F(5, 741) = 23.58$, $p <.001$. There are also important differences for properties and concepts (Table 5).

Table 5 Planned contrasts for the previoous ANOVAs (significant comparisons in bold).

	ENGLISH			DUTCH		
	Estimate	*SE*	*t*	Estimate	*SE*	*t*
(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****

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

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 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 and 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 6).

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. Louwerse and Connell (2011) showed that the peer-modalities

Table 6 Correlations among modality strength and exclusivity in the different norms (see footnote 8)

		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				–				–

*** $p < .001$. ** $p <.01$. * $p < .05$; † $p < .1$

—haptic and visual, on the one hand, and gustatory and olfactory, on the other—are peers too in the minds of comprehenders. 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 Louwerse 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 (Louwerse & Connell, 2011; Santos, Chaigneau, Simmons, & Barsalou, 2011).

Winter (submitted) 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 7).

Table 7 Examples of unimodal, bimodal, and tri-modal terms

	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, 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.

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 8 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).

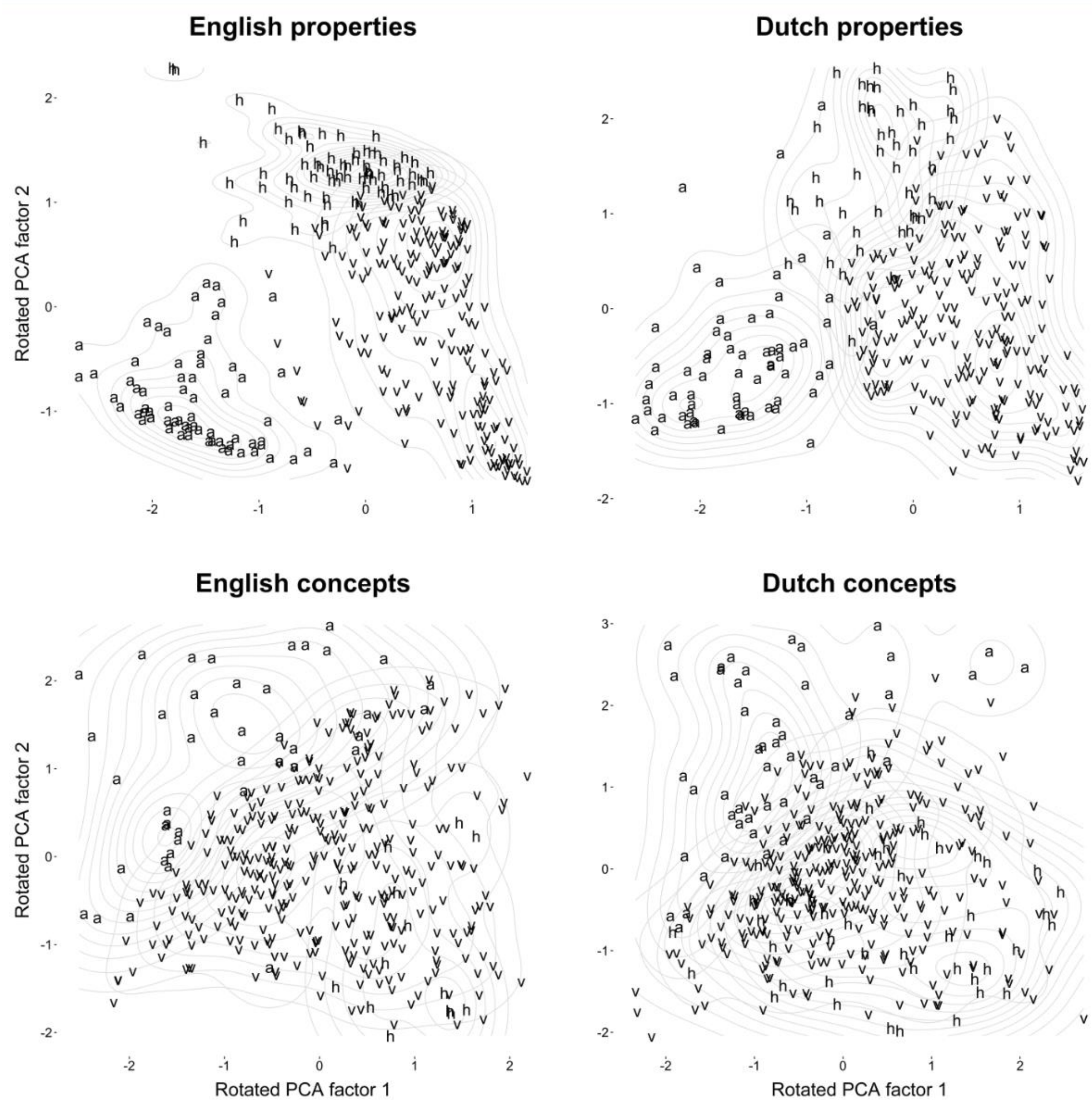
Table 8 Correlations between variables and components (no significance tested)

	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 (high-resolution figures available in the Additional Materials).

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 above plots underscore the assumption of psychometric invariance already observed 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. Further, the plots underscore the extent of the overlap between the English and the Dutch norms, the three modalities are located in very similar areas.

Sound symbolism

Sound symbolism is a 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 (Dingemanse, 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, submitted).

The same hypotheses posited by Lynott and Connell would hold for the Dutch norms (see again Table 6 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.

First, the online database CLEARPOND (Marian, Bartolotti, Chabal, & Shook, 2012) was used to estimate number of phonemes (55% properties, 95% 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 (Keuleers, Brysbaert, & New, 2010) (86% properties, 99% concepts). The CELEX lemma frquency per million was also retrieved (Baayen, Piepenbrock, & van Rijn, 1993) (74% properties, 97% concepts). In addition to those, measures for age of acquisition and concreteness were retrieved (Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014) (69% properties, 97% concepts). In regard to the properties, note that all the above measures are for neuter adjectives.[10]

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

	DUTCH PROPERTIES										
	Letters	Phonemes	Phon. Neigh.	Orth. Neigh.	Context. Diversity	Word freq.	Lemma freq.	Age of acquis.	Concrete-ness	*M*	*SD*
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

** *p* < .001. * *p* < .05

The correlations in both samples confirm the adherence of the variables within each of the three groups. In contrast to those, the extra variables AoA 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.

[10] Aside from those, the norms also contain the log-10 SUBTLEX-NL corpus diversity of inflected adjectives.

Table 11 Intercorrelations of lexical variables in the Dutch concepts sample

	DUTCH PROPERTIES										
	Letters	Phonemes	Phon. Neigh.	Orth. Neigh.	Context. Diversity	Word freq.	Lemma freq.	Age of acquis.	Concrete-ness	*M*	*SD*
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

$^{**}p < .001$. $^{*}p < .05$

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 exactly to the different groups of lexical variables (in bold).

Table 12 Correlations between variables and components (no significance tested)

	PROPERTIES			CONCEPTS		
Lexical variable	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 and Connell, 2013, Table 6). 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 ≈ 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. 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.

Table 13 Separate stepwise regressions with each lexical variable as DV and the three perceptual strengths as IVs (Dutch norms).

	PROPERTIES			CONCEPTS		
Variable	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. neighbourh.	-0.06^{**}	(+0.02)	(−0.01)	-0.04^{*}	$+0.11^{***}$	(+0.01)
Orth. neighbourh.	-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)	$+0.15^{\dagger}$	$+0.15^{**}$	(−0.05)	$+0.12^{*}$
RCs						
Length	$+0.06^{\dagger}$	(−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)	$+0.10^{\dagger}$
Other variables						
Concreteness	$-0.17^{\dagger}$	(+0.07)	$+0.18^{*}$	-0.05^{***}	$+0.11^{***}$	$+0.13^{***}$
Age of acquisition	(+0.04)	(−0.01)	-0.21^{**}	-0.10^{*}	-0.26^{***}	-0.29^{***}

Standardized (β) coefficients. Bidirectional selection, with automatic inclusion and exclusion. $N_{\text{properties}} = 336$. $N_{\text{concepts}} = 411$ (save missing lexicals). $^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$; $^{\dagger}p < .1$

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.

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

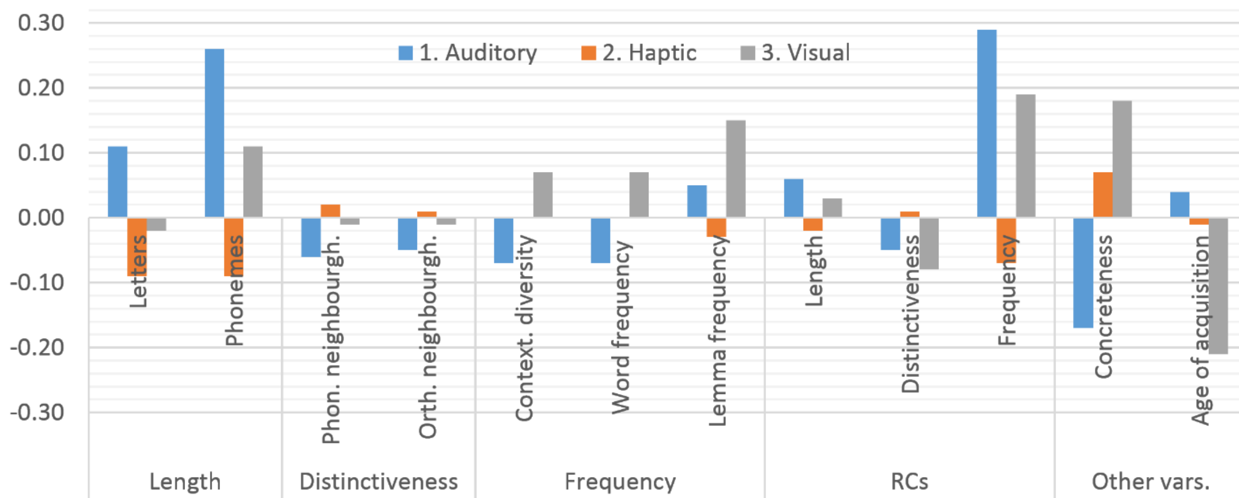
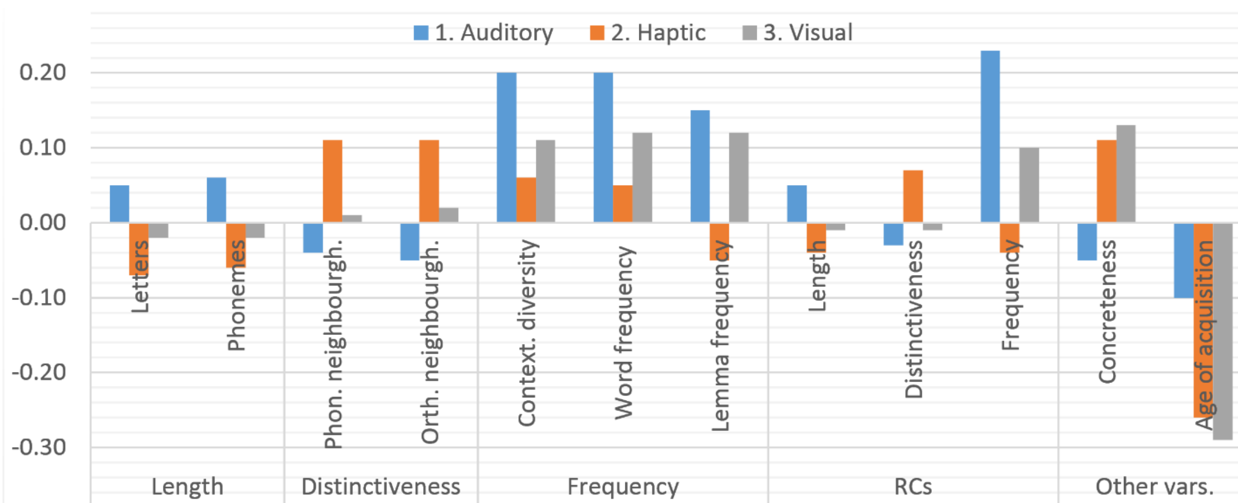


Figure 4 Graphical representation of the sound symbolism regressions for Dutch concepts



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, submitted). Nevertheless, a trend is clearly visible within Dutch properties and concepts, which furthermore converges with English words (Lynott and 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.

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 gratuitious, 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, 2016), 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.

Acknowledgments

This project would not have been possible without the generous help of Jeroen Keehnen and Wendy Leijten with the translations from English into Dutch. Equally indispensable were the forty-two respondents. Thanks also to Bodo Winter and Diane Pecher for valuable feedback.

References

Ambrosi, S., Kalénine, 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#

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*.

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*.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 3, 297-334.

Dingemanse, M., Blasi, D. E., Lupyan, 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.

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. *Front. Psychol*., *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. *Front. Psychol., 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.

Nuyts, J. (2001). *Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective.* Amsterdam: Benjamins.

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.

Newman, S. D., Klatzky, R. L., Lederman, S. J., Just, M. A. (2005). Imagining material versus geometric properties of objects: an fMRI study. *Cognit Brain Res., 23*, 235–246.

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: Modalityspecific 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.

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.

Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, *63*, 330-336.

Sutton, J., Majid, A. (2016). The content of minds. *The Psychologist, 29,* 7. Retrieved from https://thepsychologist.bps.org.uk/volume-29/july/content-minds.

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., Pecher, D., Zeelenberg, R. and Barsalou, L. W. (2008). Perceptual Processing Affects Conceptual Processing. *Cognitive Science, 32*, 579–590.

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 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). Taste and smell words form an affectively loaded and emotionally flexible part of the English lexicon, *Language, Cognition and Neuroscience*. DOI: 10.1080/23273798.2016.1193619.

Winter, B. (submitted). Sensory words are multimodal, but not extremely so.

Winter, B., & Perlman, M. (under review). Vision dominates perceptual language: Evidence from verb, noun and adjective frequencies. *Cognitive Linguistics*.

Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (submitted). Which words are most iconic? Iconicity in English sensory words. *Interaction Studies*.

Woodruff, D., & Wu, Y. (2012). Statistical considerations in choosing a test reliability coefficient. *ACT Research Report Series*, *10*.

Appendix 1 Variables in the analysis file ('all.csv'), and in the materials file ('norms.xlsx'). In parenthesis, the title of the variables left out of the materials file

Title	**Definition**
id	Item identifier. Words that are coupled across norms have the same id.
(normed)	Languages in which the word has been normed
cat	Property or concept. Applies to Dutch and English norms.
word	Word. Properties are uninflected. Variables from here up to 'word_eng' are for Dutch norms
wordcat	Linguistic word category: noun, adjective, verb, adverb
inflected_prop	Where possible, the Dutch property appears in the inflected form. Non-speakers see https://www.duolingo.com/comment/3888221
conc_cat	Indicates the inflection category of the concept: definite article 'de' or 'het'
main	Dominant modality
Exclusivity	Modality-exclusivity. Better reported as a percentage
Auditory	Mean auditory rating
Haptic	Mean haptic rating
Visual	Mean visual rating
SD_Auditory	Standard deviation of the mean auditory rating
SD_Haptic	Standard deviation of the mean haptic rating
SD_Visual	Standard deviation of the mean visual rating
freq_lg10CD_ SUBTLEXNL	Log 10 Contextual Diversity from SUBTLEX-NL corpus. See http://crr.ugent.be/programs-data/subtitle-frequencies/subtlex-nl
freq_lg10WF_ SUBTLEXNL	Log 10 Word Frequency from SUBTLEX-NL corpus. See http://crr.ugent.be/programs-data/subtitle-frequencies/subtlex-nl
freq_CELEX _lem	Lemma frequency per million, from CELEX corpus. See https://catalog.ldc.upenn.edu/LDC96L14
inflected_adj_ freq_lg10CD _SUBTLEXNL	Log 10 SUBTLEX-NL Contextual Diversity for inflected property. See http://crr.ugent.be/programs-data/subtitle-frequencies/subtlex-nl
AoA_ Brysbaertetal2014	Age of acquisition, from Brysbaert et al.'s (2014) norms. See http://crr.ugent.be/archives/1602
concrete_ Brysbaertetal2014	Concreteness, from Brysbaert et al.'s (2014) norms. See http://crr.ugent.be/archives/1602

known_ Brysbaertetal2014	Known percentage, from Brysbaert et al.'s (2014) norms. See http://crr.ugent.be/archives/1602
letters	Number of letters
phonemes_ DUTCHPOND	Number of phonemes. Retrieved from http://clearpond.northwestern.edu/dutchpond.html
orth_neighbours _DUTCHPOND	Orthographic neighbourhood size. Retrieved from http://clearpond.northwestern.edu/dutchpond.html
phon_neighbours _DUTCHPOND	Phonological neighbourhood size. Retrieved from http://clearpond.northwestern.edu/dutchpond.html
(a1)	Auditory ratings from respondent number 1 in a file
(h1)	Haptic ratings from respondent number 1 in a file
(v1)	Visual ratings from respondent number 1 in a file
(a2)	Auditory ratings from respondent number 2 in a file
(h2)	Haptic ratings from respondent number 2 in a file
(v2)	Visual ratings from respondent number 2 in a file
(a3)	Auditory ratings from respondent number 3 in a file
(h3)	Haptic ratings from respondent number 3 in a file
(v3)	Visual ratings from respondent number 3 in a file
(a4)	Auditory ratings from respondent number 4 in a file
(h4)	Haptic ratings from respondent number 4 in a file
(v4)	Visual ratings from respondent number 4 in a file
(a5)	Auditory ratings from respondent number 5 in a file
(h5)	Haptic ratings from respondent number 5 in a file
(v5)	Visual ratings from respondent number 5 in a file
(a6)	Auditory ratings from respondent number 6 in a file
(h6)	Haptic ratings from respondent number 6 in a file
(v6)	Visual ratings from respondent number 6 in a file
(a7)	Auditory ratings from respondent number 7 in a file
(h7)	Haptic ratings from respondent number 7 in a file
(v7)	Visual ratings from respondent number 7 in a file
(a8)	Auditory ratings from respondent number 8 in a file
(h8)	Haptic ratings from respondent number 8 in a file
(v8)	Visual ratings from respondent number 8 in a file

(a9)	Auditory ratings from respondent number 9 in a file
(h9)	Haptic ratings from respondent number 9 in a file
(v9)	Visual ratings from respondent number 9 in a file
(a10)	Auditory ratings from respondent number 10 in a file
(h10)	Haptic ratings from respondent number 10 in a file
(v10)	Visual ratings from respondent number 10 in a file
(file)	Identifier of the file(s) in which the word was rated
(word_eng)	English word. All variables hereafter for these words (Lynott and Connell, 2009, 2013). Retrieved from http://www.lancaster.ac.uk/people/connelll/lab/norms.html.
(main_eng)	Dominant modality
(exc_eng)	Modality exclusivity
(Aud_eng)	Mean auditory rating
(Hap_eng)	Mean haptic rating
(Vis_eng)	Mean visual rating
(lett_eng)	Number of letters

Appendix 2 Excerpt from the analysis file ('all.csv'). This file was put together in Excel. For analysis purposes, it includes the individual ratings from each respondent, as well as the English means from Lynott and Connell (2009, 2013). For users' convenience, however, the materials file ('norms.xlsx') does not include the individual ratings or the English ratings.

normed	cat	word	main	Exclusivity	word_eng	main_eng	exc_eng
Dutch	conc	aankondiging	a	0.48	announcement		
Dut_Eng	conc	aantekening	v	0.22	note	v	0.46
Dut_Eng	conc	aanvraag	v	0.32	appeal	v	0.22
Dut_Eng	prop	aards	v	0.14	earthy	v	0.21
Dut_Eng	prop	absorberend	v	0.27	absorbent	v	0.41
Dut_Eng	conc	academie	v	0.36	academy	v	0.56
Dut_Eng	conc	achtergrond	v	0.51	background	v	0.53
Dut_Eng	conc	administratie	v	0.39	administration	v	0.49
Dut_Eng	conc	advies	a	0.41	advice	a	0.70
Dut_Eng	conc	afbeelding	v	0.58	picture	v	0.68
Dut_Eng	conc	afdeling	v	0.57	department	v	0.51
Dut_Eng	conc	afsluiting	v	0.35	close	v	0.36
Dut_Eng	conc	afstand	v	0.46	distance	v	0.50
Dut_Eng	conc	afval	v	0.34	waste	v	0.30
Dut_Eng	conc	amateur	v	0.23	amateur	v	0.42
Dut_Eng	conc	ambacht	v	0.33	craft	v	0.33
Dut_Eng	prop	amber	v	0.42	amber	v	0.66
Dut_Eng	conc	angst	v	0.14	fear	v	0.31
Dut_Eng	conc	antwoord	a	0.42	answer	a	0.55
Dut_Eng	conc	apparaat	v	0.10	device	v	0.38
Dut_Eng	conc	arbeider	v	0.23	worker	v	0.44
Dut_Eng	conc	arrangement	v	0.46	arrangement	v	0.42
Dut_Eng	conc	aspect	v	0.38	aspect	v	0.29
Dut_Eng	conc	atoom	v	0.29	atom	v	0.20
Dut_Eng	conc	baan	v	0.24	job	v	0.33
Dut_Eng	conc	baas	v	0.49	boss	v	0.42
Dut_Eng	conc	baby	v	0.04	baby	v	0.24

Dut_Eng	conc	bad	h	0.16	bath	h	0.30
Dut_Eng	conc	balans	v	0.16	balance	v	0.40
Dut_Eng	conc	band	v	0.33	band	a	0.46
Dut_Eng	conc	bank	v	0.42	bank	v	0.60
Dut_Eng	prop	barstend	v	0.11	bursting	v	0.26
Dut_Eng	conc	basis	v	0.17	base	v	0.46
Dut_Eng	conc	bedrag	v	0.23	amount	v	0.22
Dut_Eng	conc	been	v	0.28	leg	v	0.41
Dut_Eng	conc	beer	v	0.32	bear	v	0.32
Dut_Eng	conc	begin	v	0.18	beginning	v	0.30
Dut_Eng	prop	behaard	v	0.44	hairy	v	0.45
Dut_Eng	conc	beheer	v	0.40	management	v	0.48
Dut_Eng	conc	behoefte	h	0.09	want	v	0.12
Dut_Eng	prop	beige	v	0.54	beige	v	0.92
Dut_Eng	conc	beker	v	0.47	cup	v	0.27
Dut_Eng	conc	belading	v	0.26	load	v	0.36
Dut_Eng	conc	belangrijkste	v	0.26	main	v	0.37
Dut_Eng	conc	belasting	v	0.20	tax	v	0.61
Dut_Eng	conc	beneden	v	0.53	down	v	0.54
Dut_Eng	prop	beschimmeld	v	0.50	mouldy	v	0.34
Dutch	conc	besmetting	h	0.15	contagion		
Dut_Eng	conc	bestand	v	0.48	file	v	0.52
Dut_Eng	prop	betoverend	v	0.26	glamorous	v	0.47
Dut_Eng	conc	beurs	v	0.05	fair	v	0.48
Dut_Eng	conc	beurt	v	0.22	turn	v	0.50
Dut_Eng	prop	bevriezend	h	0.38	freezing	h	0.34
Dut_Eng	conc	bewijs	v	0.27	proof	v	0.24
Dut_Eng	prop	bewolkt	v	0.53	cloudy	v	0.81
Dut_Eng	conc	bezit	v	0.10	estate	v	0.53
Dut_Eng	conc	bibliotheek	v	0.18	library	v	0.33
Dut_Eng	prop	blaffend	a	0.47	barking	a	0.51
Dut_Eng	prop	blauw	v	0.68	blue	v	0.80
Dut_Eng	prop	bleek	v	0.62	pale	v	0.82

Dut_Eng	prop	blij	v	0.20	happy	v	0.27
Dut_Eng	conc	blijven	v	0.24	stay	v	0.38
Dut_Eng	conc	blik	v	0.45	look	v	0.68
Dut_Eng	prop	blinkend	v	0.53	shiny	v	0.70
Dut_Eng	prop	bloedig	v	0.35	bloody	v	0.41
Dut_Eng	prop	bloemrijk	v	0.40	flowery	v	0.41
Dut_Eng	prop	blond	v	0.68	blonde	v	0.91
Dut_Eng	conc	bocht	v	0.53	curve	v	0.51
Dut_Eng	conc	boerderij	v	0.19	farm	v	0.25
Dut_Eng	conc	boete	v	0.31	fine	v	0.41
Dut_Eng	prop	bolvormig	v	0.49	globular	v	0.43
Dut_Eng	prop	bonzend	a	0.32	thudding	a	0.46
Dut_Eng	conc	boosheid	v	0.31	anger	v	0.41
Dut_Eng	prop	borstelig	h	0.34	bristly	h	0.37
Dut_Eng	conc	bot	h	0.19	bone	h	0.27
Dut_Eng	prop	botsend	a	0.07	crashing	v	0.40
Dut_Eng	conc	bouw	v	0.33	construction	v	0.39
Dut_Eng	conc	bouwer	v	0.40	builder	v	0.38
Dut_Eng	prop	brak	v	0.20	brackish	h	0.15
Dut_Eng	prop	breed	v	0.45	wide	v	0.50
Dut_Eng	prop	breekbaar	v	0.16	breakable	v	0.39
Dut_Eng	conc	breuk	h	0.22	break	v	0.25
Dut_Eng	prop	briesend	a	0.07	snorting	a	0.51
Dut_Eng	prop	briljant	v	0.25	brilliant	v	0.13
Dut_Eng	prop	brommend	a	0.29	snarling	a	0.54
Dut_Eng	prop	bronzen	v	0.77	bronze	v	0.68
Dut_Eng	prop	broos	h	0.30	brittle	h	0.42
Dut_Eng	prop	bruin	v	0.62	brown	v	0.83
Dut_Eng	prop	bruinharig	v	0.40	brunette	v	0.98
Dut_Eng	conc	bureau	v	0.45	desk	v	0.43
Dut_Eng	conc	capaciteit	v	0.14	capacity	v	0.37
Dut_Eng	conc	carrière	v	0.25	career	v	0.39
Dut_Eng	conc	centrum	v	0.60	center	v	0.47

Dut_Eng	conc	cijfer	v	0.38	grade	v	0.61
Dut_Eng	prop	circulair	a	0.06	circular	v	0.54
Dut_Eng	prop	compact	v	0.38	compact	v	0.53
Dutch	conc	compliment	a	0.39	compliment		
Dut_Eng	conc	concept	v	0.34	concept	v	0.23
Dut_Eng	conc	concurrentie	v	0.18	competition	v	0.37
Dut_Eng	conc	conditie	v	0.15	condition	v	0.24
Dut_Eng	prop	conisch	v	0.32	conical	v	0.52
Dut_Eng	conc	consequentie	v	0.19	consequence	v	0.31
Dut_Eng	conc	contact	h	0.14	contact	h	0.21
Dut_Eng	conc	contract	v	0.27	contract	v	0.46
Dut_Eng	conc	crisis	v	0.13	crisis	v	0.41
Dut_Eng	conc	dak	v	0.48	roof	v	0.61
Dut_Eng	conc	dame	v	0.18	lady	v	0.27
Dut_Eng	conc	dans	v	0.21	dance	v	0.50
Dut_Eng	conc	deel	v	0.55	portion	v	0.33
Dut_Eng	prop	deftig	v	0.40	portly	v	0.44
Dut_Eng	conc	dek	v	0.30	deck	v	0.47
Dut_Eng	conc	deken	h	0.28	blanket	h	0.39
Dut_Eng	conc	democratie	v	0.24	democracy	v	0.53
Dut_Eng	conc	depressie	v	0.36	depression	v	0.43
Dut_Eng	conc	diameter	v	0.68	bore	v	0.31
Dut_Eng	conc	dichter	a	0.22	poet	a	0.47
Dut_Eng	conc	dichterbij	v	0.24	closer	v	0.28
Dut_Eng	prop	diep	v	0.25	deep	v	0.38
Dut_Eng	prop	dik	v	0.34	fat	v	0.38
Dut_Eng	conc	dik	v	0.53	thick	v	0.30
Dut_Eng	conc	doel	a	0.21	purpose	v	0.40
Dut_Eng	prop	dof	v	0.28	dull	v	0.39
Dut_Eng	conc	dokter	v	0.08	doctor	v	0.34
Dut_Eng	prop	donderend	a	0.36	thunderous	a	0.54
Dut_Eng	prop	donker	v	0.63	dark	v	0.70
Dut_Eng	prop	donzig	h	0.26	downy	v	0.52

Dut_Eng	prop	dood	v	0.53	dead	v	0.37
Dut_Eng	conc	dood	v	0.32	death	v	0.37
Dut_Eng	prop	doornig	v	0.37	thorny	h	0.42
Dut_Eng	prop	doorschijnend	v	0.70	translucent	v	0.68
Dut_Eng	prop	doorweekt	h	0.30	sodden	v	0.42
Dut_Eng	conc	drama	v	0.26	drama	v	0.42
Dut_Eng	prop	drapperig	h	0.19	slushy	v	0.27
Dut_Eng	prop	drassig	h	0.16	soggy	h	0.24
Dut_Eng	conc	drie	v	0.63	three	v	0.39
Dut_Eng	prop	driehoekig	v	0.51	triangular	v	0.54
Dut_Eng	prop	droog	v	0.46	dry	h	0.35
Dut_Eng	conc	droom	v	0.24	dream	v	0.35
Dut_Eng	prop	druk	v	0.22	crowded	v	0.41
Dut_Eng	conc	druk	v	0.27	pressure	v	0.36
Dut_Eng	conc	drukte	v	0.20	rush	v	0.38
Dut_Eng	conc	duik	h	0.11	dive	v	0.45
Dut_Eng	conc	duim	v	0.41	inch	v	0.60
Dut_Eng	prop	echoënd	a	0.80	echoing	a	0.85
Dut_Eng	conc	economie	a	0.35	economy	v	0.44
Dut_Eng	conc	eenheid	v	0.32	unit	v	0.37
Dut_Eng	conc	eeuw	v	0.91	century	v	0.50
Dut_Eng	conc	effect	v	0.41	effect	v	0.19
Dut_Eng	prop	effen	v	0.46	plain	v	0.36
Dut_Eng	conc	eigenaar	v	0.24	owner	v	0.46
Dut_Eng	conc	eigenschap	v	0.10	property	v	0.44
Dut_Eng	conc	eis	a	0.33	requirement	v	0.42
Dut_Eng	prop	elastisch	v	0.25	elastic	h	0.34
Dut_Eng	prop	elegant	v	0.50	elegant	v	0.41
Dut_Eng	conc	emotie	v	0.14	emotion	v	0.27
Dut_Eng	prop	enorm	v	0.28	enormous	v	0.46
Dut_Eng	conc	enthousiasme	a	0.08	enthusiasm	v	0.34
Dut_Eng	conc	entree	v	0.53	entrance	v	0.52
Dut_Eng	conc	erkenning	a	0.18	recognition	v	0.15

Dut_Eng	conc	exemplaar	v	0.27	instance	v	0.39
Dut_Eng	conc	expansie	v	0.25	expansion	v	0.38
Dut_Eng	conc	extreem	v	0.14	extreme	v	0.17
Dut_Eng	conc	fabriek	v	0.41	factory	v	0.31
Dut_Eng	conc	factor	v	0.09	factor	v	0.46
Dut_Eng	conc	feit	a	0.30	fact	a	0.28
Dutch	conc	felicitatie	a	0.28	congratulation		
Dut_Eng	conc	filosofie	a	0.19	philosophy	a	0.64
Dut_Eng	conc	financiën	v	0.39	finance	v	0.48
Dut_Eng	conc	firma	v	0.13	firm	h	0.42
Dut_Eng	conc	flauwvallen	v	0.24	faint	v	0.40
Dut_Eng	prop	flikkerend	v	0.54	flickering	v	0.69
Dut_Eng	prop	floraal	v	0.73	floral	v	0.40
Dut_Eng	prop	fluisterend	a	0.39	whispering	a	0.62
Dut_Eng	prop	fluitend	a	0.30	bleeping	a	0.69
Dut_Eng	prop	fonkelend	v	0.38	glistening	v	0.67
Dut_Eng	conc	formatie	v	0.38	formation	v	0.48
Dut_Eng	conc	fortuin	v	0.25	fortune	v	0.43
Dut_Eng	conc	fout	v	0.10	wrong	v	0.20
Dut_Eng	prop	fuchsia	v	0.92	tangerine	v	0.29
Dut_Eng	conc	functie	v	0.15	function	v	0.30
Dut_Eng	prop	galmend	a	0.74	resounding	a	0.61
Dut_Eng	conc	gat	v	0.33	hole	v	0.54
Dut_Eng	conc	gebaar	v	0.59	gesture	v	0.60
Dut_Eng	conc	gebied	v	0.35	area	v	0.50
Dut_Eng	prop	geblokt	v	0.42	chequered	v	0.92
Dut_Eng	prop	gebogen	v	0.27	bent	v	0.53
Dut_Eng	prop	gebroken	h	0.11	broken	v	0.33
Dut_Eng	prop	gedempt	a	0.68	muffled	a	0.60
Dut_Eng	prop	geel	v	0.67	yellow	v	0.95
Dut_Eng	conc	geest	h	0.24	spirit	v	0.35
Dut_Eng	conc	gelach	a	0.33	laughter	a	0.49
Dut_Eng	conc	geld	v	0.17	cash	v	0.37

Dut_Eng	conc	gelegenheid	v	0.50	opportunity	v	0.37
Dut_Eng	conc	geluid	a	0.57	sound	a	0.78
Dut_Eng	prop	geluidloos	a	0.52	soundless	a	0.67
Dut_Eng	conc	gemak	v	0.12	ease	v	0.41
Dut_Eng	prop	geneeskrachtig	h	0.18	medicinal	v	0.28
Dut_Eng	conc	genezing	v	0.14	cure	v	0.32
Dut_Eng	conc	genot	v	0.03	delight	v	0.16
Dut_Eng	prop	gepatroneerd	v	0.32	patterned	v	0.68
Dut_Eng	prop	geplooid	v	0.16	crinkled	v	0.33
Dut_Eng	prop	gepolijst	v	0.42	polished	v	0.45
Dut_Eng	conc	gereedschap	v	0.24	tool	v	0.36
Dutch	conc	gerucht	a	0.46	rumour		
Dut_Eng	conc	geschil	v	0.20	dispute	v	0.44
Dut_Eng	prop	geschubd	h	0.28	scaly	h	0.43
Dut_Eng	prop	gesmolten	v	0.47	melted	v	0.29
Dut_Eng	prop	gespikkeld	v	0.85	speckled	v	0.67
Dut_Eng	prop	gestreept	v	0.64	striped	v	0.76
Dut_Eng	prop	getand	v	0.29	jagged	h	0.42
Dut_Eng	prop	gevlekt	v	0.68	dappled	v	0.55
Dut_Eng	conc	gevoel	v	0.12	feel	h	0.41
Dut_Eng	prop	gevorkt	v	0.37	forked	v	0.49
Dut_Eng	prop	gevormd	v	0.15	contoured	v	0.50
Dut_Eng	prop	gevouwen	v	0.49	creased	v	0.52
Dut_Eng	conc	geweer	h	0.09	rifle	a	0.31
Dut_Eng	prop	gewelfd	v	0.26	curved	v	0.52
Dut_Eng	prop	gewichtloos	h	0.26	weightless	h	0.46
Dut_Eng	prop	gezet	v	0.45	rotund	v	0.57
Dut_Eng	conc	gezondheid	v	0.10	health	v	0.32
Dut_Eng	prop	gezwollen	v	0.41	puffy	v	0.39
Dut_Eng	prop	giechelend	a	0.39	giggling	a	0.51
Dut_Eng	prop	gierend	a	0.50	howling	a	0.59
Dut_Eng	prop	gigantisch	v	0.73	gigantic	v	0.47
Dut_Eng	prop	gillend	a	0.33	screaming	a	0.61

Dut_Eng	conc	gips	v	0.35	cast	v	0.47
Dut_Eng	prop	glad	h	0.33	slick	v	0.40
Dut_Eng	prop	glanzend	v	0.78	glossy	v	0.46
Dut_Eng	conc	glas	v	0.15	glass	v	0.39
Dut_Eng	prop	glibberig	h	0.36	slippery	h	0.35
Dut_Eng	prop	glimmend	v	0.58	gleaming	v	0.65
Dut_Eng	prop	glinsterend	v	0.68	shimmering	v	0.65
Dut_Eng	prop	glitterend	v	0.52	glittery	v	0.61
Dut_Eng	prop	gloeiend	v	0.50	glowing	v	0.74
Dut_Eng	conc	god	v	0.11	god	a	0.33
Dut_Eng	prop	golvend	v	0.17	rippled	v	0.37
Dut_Eng	prop	gorgelend	a	0.29	gurgling	a	0.42
Dut_Eng	prop	gouden	v	0.56	gold	v	0.72
Dut_Eng	conc	grafiek	v	0.50	chart	v	0.74
Dut_Eng	conc	gras	v	0.44	grass	v	0.35
Dut_Eng	prop	grasachtig	h	0.32	grassy	v	0.39
Dut_Eng	conc	grens	v	0.29	border	v	0.58
Dut_Eng	prop	grijs	v	0.65	grey	v	0.85
Dut_Eng	prop	groeiend	v	0.43	booming	a	0.57
Dut_Eng	prop	groen	v	0.58	green	v	0.69
Dut_Eng	prop	grof	v	0.25	harsh	a	0.12
Dut_Eng	prop	grommend	a	0.48	growling	a	0.56
Dut_Eng	prop	groot	v	0.38	big	v	0.36
Dut_Eng	prop	grotesk	v	0.86	grotesque	v	0.18
Dut_Eng	conc	haat	v	0.11	hate	a	0.31
Dut_Eng	prop	hard	h	0.14	hard	h	0.42
Dut_Eng	conc	harmonie	v	0.25	harmony	a	0.34
Dut_Eng	prop	hees	a	0.71	hoarse	a	0.51
Dut_Eng	prop	heet	h	0.40	hot	h	0.25
Dutch	conc	heft	h	0.33	handle		
Dut_Eng	conc	hek	v	0.36	fence	v	0.52
Dut_Eng	conc	hel	h	0.19	hell	a	0.39
Dut_Eng	prop	helder	v	0.54	bright	v	0.90

Dut_Eng	conc	hemel	v	0.42	heaven	a	0.21
Dut_Eng	conc	herinnering	a	0.25	recall	v	0.14
Dut_Eng	conc	hinder	v	0.18	bother	v	0.34
Dut_Eng	prop	hitsig	h	0.13	searing	v	0.36
Dut_Eng	prop	hobbelig	v	0.32	bumpy	h	0.42
Dut_Eng	conc	hoed	v	0.47	hat	v	0.52
Dut_Eng	prop	hoekig	v	0.31	angular	v	0.49
Dut_Eng	conc	hoeveelheid	v	0.30	quantity	v	0.40
Dut_Eng	prop	hol	v	0.24	hollow	v	0.33
Dut_Eng	prop	hoog	v	0.69	high	v	0.42
Dut_Eng	prop	hoorbaar	a	0.77	audible	a	0.76
Dut_Eng	prop	huilend	v	0.30	crying	a	0.43
Dut_Eng	prop	huilerig	v	0.17	wailing	a	0.58
Dut_Eng	conc	huishouden	v	0.17	household	v	0.22
Dut_Eng	conc	hulp	v	0.04	help	a	0.37
Dut_Eng	conc	huur	v	0.32	rent	v	0.50
Dut_Eng	conc	ideaal	v	0.20	ideal	v	0.18
Dut_Eng	prop	iel	v	0.26	puny	v	0.49
Dut_Eng	prop	ijsachtig	v	0.35	icy	v	0.31
Dut_Eng	prop	ijzig	v	0.17	frosty	v	0.30
Dut_Eng	conc	incident	v	0.28	incident	v	0.35
Dut_Eng	conc	indruk	v	0.15	impression	v	0.29
Dut_Eng	conc	informatie	v	0.29	information	a	0.27
Dut_Eng	conc	inhoud	v	0.37	content	v	0.20
Dut_Eng	conc	inkomen	v	0.23	income	v	0.52
Dut_Eng	conc	inspanning	v	0.15	effort	v	0.45
Dut_Eng	conc	interieur	v	0.52	interior	v	0.44
Dut_Eng	conc	invloed	v	0.21	influence	v	0.37
Dut_Eng	conc	item	v	0.31	item	h	0.08
Dut_Eng	prop	jammerend	a	0.25	whimpering	a	0.56
Dut_Eng	prop	jankend	v	0.21	squealing	a	0.68
Dut_Eng	prop	jeukend	h	0.33	itchy	h	0.46
Dut_Eng	prop	jodelend	a	0.28	warbling	a	0.51

Dut_Eng	conc	juffrouw	v	0.25	miss	v	0.34
Dut_Eng	conc	junior	v	0.14	junior	v	0.38
Dut_Eng	prop	kaatsend	v	0.24	bouncy	h	0.40
Dut_Eng	prop	kabbelend	v	0.19	rippling	v	0.34
Dut_Eng	conc	kamp	v	0.25	camp	v	0.30
Dut_Eng	prop	karmozijn	v	0.75	crimson	v	0.87
Dut_Eng	conc	katoen	h	0.51	cotton	h	0.44
Dut_Eng	conc	kelder	v	0.41	cellar	v	0.36
Dut_Eng	conc	kennis	a	0.32	knowledge	a	0.46
Dut_Eng	conc	kern	h	0.18	core	v	0.35
Dut_Eng	conc	keten	v	0.21	chain	v	0.38
Dut_Eng	prop	kietelend	h	0.31	ticklish	h	0.52
Dut_Eng	prop	kil	v	0.29	chilly	h	0.37
Dut_Eng	conc	kind	v	0.10	kid	v	0.38
Dut_Eng	conc	kip	v	0.06	hen	v	0.22
Dut_Eng	prop	kladderig	v	0.54	blotchy	v	0.60
Dut_Eng	prop	klam	h	0.47	clammy	h	0.41
Dut_Eng	conc	klant	v	0.23	customer	v	0.37
Dutch	conc	klap	h	0.09	slap		
Dut_Eng	prop	klappend	a	0.08	banging	a	0.51
Dut_Eng	prop	klein	v	0.48	small	v	0.55
Dut_Eng	prop	kleurrijk	v	0.65	colourful	v	0.77
Dut_Eng	prop	kleverig	h	0.32	sticky	h	0.43
Dut_Eng	prop	klonterig	v	0.28	lumpy	h	0.34
Dut_Eng	prop	knallend	a	0.37	popping	a	0.36
Dut_Eng	prop	knap	v	0.52	handsome	v	0.51
Dut_Eng	prop	knapperend	a	0.27	crackling	a	0.32
Dut_Eng	prop	knapperig	h	0.00	crisp	h	0.18
Dut_Eng	prop	knarsend	a	0.32	creaking	a	0.58
Dutch	conc	knuffel	h	0.44	hug		
Dut_Eng	prop	koel	h	0.46	cool	h	0.44
Dut_Eng	conc	koelkast	v	0.23	refrigerator	v	0.26
Dut_Eng	prop	koerend	a	0.65	cooing	a	0.68

Dut_Eng	conc	kogel	v	0.15	bullet	v	0.32
Dut_Eng	prop	kokend	v	0.14	boiling	v	0.24
Dut_Eng	prop	kolossaal	v	0.29	colossal	v	0.47
Dut_Eng	conc	komedie	a	0.22	comedy	a	0.51
Dut_Eng	conc	koninkrijk	v	0.34	kingdom	v	0.39
Dut_Eng	prop	koortsig	v	0.25	feverish	h	0.41
Dut_Eng	prop	korrelig	h	0.34	grainy	h	0.32
Dut_Eng	prop	kort	v	0.58	short	v	0.53
Dut_Eng	conc	kosten	v	0.31	cost	v	0.45
Dut_Eng	prop	koud	h	0.47	cold	h	0.34
Dut_Eng	prop	krakend	a	0.48	crunching	a	0.29
Dut_Eng	prop	krassend	a	0.18	scratchy	h	0.39
Dut_Eng	prop	kreunend	a	0.50	groaning	a	0.66
Dut_Eng	prop	krijsend	a	0.23	screeching	a	0.61
Dut_Eng	prop	krullend	v	0.22	curly	v	0.53
Dutch	conc	kus	h	0.18	kiss		
Dut_Eng	conc	kwaliteit	v	0.12	quality	v	0.07
Dut_Eng	conc	kwestie	v	0.24	matter	v	0.17
Dut_Eng	prop	laag	v	0.63	low	v	0.46
Dut_Eng	prop	lachend	a	0.50	laughing	a	0.48
Dut_Eng	conc	landschap	v	0.34	landscape	v	0.55
Dut_Eng	prop	lang	v	0.53	long	v	0.51
Dut_Eng	prop	langgerekt	v	0.38	tall	v	0.63
Dutch	conc	laster	a	0.25	libel		
Dut_Eng	prop	lauw	a	0.66	lukewarm	h	0.50
Dut_Eng	prop	leeg	v	0.44	empty	v	0.44
Dut_Eng	prop	leerachtig	v	0.24	leathery	h	0.34
Dut_Eng	conc	leiding	a	0.13	lead	v	0.33
Dut_Eng	prop	lelijk	v	0.47	ugly	v	0.46
Dut_Eng	prop	lenig	v	0.50	lithe	v	0.56
Dutch	conc	letsel	v	0.38	injury		
Dut_Eng	conc	leven	v	0.06	life	v	0.08
Dut_Eng	prop	levend	v	0.10	alive	v	0.18

Dut_Eng	prop	levendig	v	0.28	vivid	v	0.29
Dut_Eng	conc	levering	v	0.05	supply	v	0.40
Dut_Eng	conc	lichaam	v	0.27	body	v	0.19
Dut_Eng	prop	licht	v	0.51	faint	v	0.24
Dut_Eng	prop	lichtbruin	v	0.70	khaki	v	0.79
Dut_Eng	conc	lift	v	0.15	lift	v	0.39
Dut_Eng	conc	links	v	0.28	left	v	0.48
Dut_Eng	conc	loon	v	0.16	wage	v	0.53
Dut_Eng	prop	los	v	0.58	loose	v	0.48
Dut_Eng	prop	luchtig	v	0.38	breezy	v	0.32
Dut_Eng	prop	luid	a	0.63	loud	a	0.68
Dut_Eng	prop	luidruchtig	a	0.53	noisy	a	0.70
Dut_Eng	conc	lus	v	0.56	loop	v	0.43
Dut_Eng	conc	maat	v	0.28	measure	v	0.32
Dut_Eng	conc	maatje	v	0.09	mate	v	0.33
Dut_Eng	conc	machine	v	0.21	machine	v	0.32
Dut_Eng	prop	mager	v	0.47	skinny	v	0.50
Dut_Eng	conc	magie	v	0.48	magic	v	0.51
Dut_Eng	prop	mals	h	0.35	tender	h	0.22
Dut_Eng	conc	mandaat	v	0.12	brief	a	0.34
Dutch	conc	manipulatie	v	0.05	manipulation		
Dut_Eng	conc	mars	v	0.14	march	v	0.42
Dut_Eng	conc	massa	v	0.22	mass	v	0.43
Dutch	conc	massage	h	0.38	massage		
Dut_Eng	conc	meerderheid	v	0.42	majority	v	0.50
Dut_Eng	conc	meester	v	0.41	master	v	0.41
Dut_Eng	prop	melodieus	a	0.67	melodious	a	0.64
Dut_Eng	prop	metalen	v	0.20	metal	v	0.24
Dut_Eng	conc	methode	v	0.33	method	v	0.43
Dut_Eng	prop	miauwend	a	0.54	meowing	a	0.59
Dut_Eng	prop	miniatuur	v	0.49	miniature	v	0.51
Dut_Eng	conc	minuut	v	0.34	minute	v	0.54
Dut_Eng	conc	misdrijf	v	0.22	crime	v	0.45

Dut_Eng	conc	mislukking	v	0.18	failure	v	0.31
Dut_Eng	conc	mist	v	0.43	fog	v	0.66
Dut_Eng	prop	mistig	v	0.46	foggy	v	0.72
Dut_Eng	prop	modderig	v	0.44	muddy	v	0.30
Dut_Eng	prop	mollig	v	0.31	chubby	v	0.47
Dut_Eng	conc	moment	v	0.24	moment	v	0.33
Dut_Eng	prop	mompelend	a	0.61	mumbling	a	0.66
Dut_Eng	conc	mond	v	0.05	mouth	v	0.17
Dut_Eng	prop	mondig	a	0.53	mellow	v	0.33
Dut_Eng	prop	mooi	v	0.18	beautiful	v	0.12
Dut_Eng	prop	morrend	a	0.53	muttering	a	0.60
Dut_Eng	prop	morsend	v	0.29	spilling	v	0.45
Dut_Eng	prop	mottig	v	0.14	dank	v	0.30
Dut_Eng	conc	muziek	a	0.47	music	a	0.58
Dut_Eng	conc	mythe	a	0.28	myth	a	0.73
English	prop				whistling	a	0.65
English	prop				tepid	h	0.49
English	prop				spotted	v	0.68
English	prop				clear	v	0.40
English	prop				mottled	v	0.48
English	prop				light	v	0.31
English	prop				broad	v	0.45
English	conc				blame	a	0.38
English	conc				claim	a	0.50
English	conc				aim	v	0.41
English	conc				confidence	v	0.39
English	conc				money	v	0.39
Dut_Eng	conc	nacht	v	0.45	night	v	0.62
Dut_Eng	prop	nassig	v	1.00	damp	h	0.32
Dut_Eng	prop	nat	h	0.27	wet	h	0.21
Dut_Eng	prop	nattig	v	0.14	moist	h	0.21
Dut_Eng	prop	nauw	v	0.33	narrow	v	0.55
Dut_Eng	conc	nederlaag	v	0.09	defeat	v	0.46

Dut_Eng	conc	neef	v	0.15	cousin	v	0.34
Dut_Eng	conc	neiging	a	0.19	tendency	v	0.47
Dut_Eng	conc	neus	h	0.22	nose	v	0.30
Dut_Eng	prop	nevelig	v	0.48	misty	v	0.54
Dut_Eng	conc	nieuwsgierigheid	v	0.15	curiosity	v	0.22
Dut_Eng	conc	niveau	v	0.48	level	v	0.44
Dut_Eng	conc	noodgeval	a	0.25	emergency	v	0.38
Dut_Eng	conc	nul	v	0.59	zero	v	0.55
Dut_Eng	conc	object	v	0.26	object	v	0.13
Dut_Eng	prop	olieachtig	v	0.47	oily	v	0.28
Dut_Eng	conc	omzoming	v	0.66	trim	v	0.46
Dut_Eng	conc	onafhankelijkheid	h	0.14	independence	v	0.43
Dut_Eng	prop	onbeweeglijk	v	0.44	motionless	v	0.52
Dut_Eng	conc	onder	v	0.47	under	v	0.51
Dut_Eng	conc	onderscheid	v	0.19	distinction	v	0.15
Dut_Eng	conc	onderwijs	v	0.39	education	v	0.35
Dut_Eng	conc	onderzoek	v	0.14	investigation	v	0.41
Dut_Eng	conc	ondeugd	v	0.16	vice	v	0.38
Dut_Eng	prop	ondiep	v	0.42	shallow	v	0.43
Dut_Eng	prop	ongelijk	v	0.25	uneven	v	0.37
Dut_Eng	prop	onmetelijk	v	0.48	immense	v	0.20
Dut_Eng	prop	onrijp	v	0.43	unripe	v	0.27
Dutch	conc	ontharder	v	0.42	softener		
Dut_Eng	conc	ontwikkeling	v	0.27	development	v	0.27
Dut_Eng	conc	oog	v	0.66	eye	v	0.68
Dut_Eng	conc	oorsprong	v	0.06	origin	v	0.36
Dut_Eng	prop	oorverdovend	a	0.67	deafening	a	0.77
Dut_Eng	conc	oorzaak	v	0.31	cause	v	0.36
Dut_Eng	conc	oosten	v	0.64	east	v	0.63
Dut_Eng	prop	open	v	0.33	open	v	0.48
Dut_Eng	conc	openbaar	v	0.33	public	v	0.39
Dut_Eng	conc	opening	v	0.14	opening	v	0.42
Dut_Eng	conc	oplossing	v	0.31	solution	v	0.38

Dut_Eng	conc	opwinding	v	0.01	excitement	v	0.23
Dut_Eng	prop	oranje	v	0.66	orange	v	0.41
Dut_Eng	prop	ovaal	v	0.51	oval	v	0.50
Dut_Eng	conc	overeenkomst	a	0.15	deal	v	0.35
Dut_Eng	conc	overhemd	v	0.37	shirt	v	0.42
Dut_Eng	conc	overwinning	a	0.07	win	v	0.38
Dut_Eng	conc	paar	v	0.48	pair	v	0.50
Dut_Eng	prop	paars	v	0.79	purple	v	0.90
Dutch	conc	palpatie	h	0.42	palpation		
Dut_Eng	conc	papier	v	0.27	paper	v	0.32
Dut_Eng	prop	papperig	v	0.22	mushy	h	0.28
Dut_Eng	conc	pels	v	0.36	hide	v	0.50
Dut_Eng	conc	personeel	v	0.38	personnel	v	0.47
Dut_Eng	prop	piekerig	v	0.48	wispy	v	0.38
Dut_Eng	prop	piepend	a	0.42	squeaking	a	0.65
Dut_Eng	prop	piepklein	v	0.63	tiny	v	0.52
Dut_Eng	prop	pijnlijk	v	0.05	painful	h	0.29
Dut_Eng	prop	pijnvol	h	0.08	aching	h	0.55
Dut_Eng	conc	pistool	h	0.09	gun	v	0.29
Dut_Eng	conc	plaats	v	0.44	site	v	0.53
Dut_Eng	prop	plakkend	h	0.24	adhesive	h	0.29
Dut_Eng	conc	plan	a	0.26	plan	a	0.42
Dut_Eng	prop	plantaardig	v	0.35	vegetal	v	0.33
Dut_Eng	prop	plastic	v	0.23	plastic	v	0.27
Dut_Eng	conc	platform	v	0.39	platform	v	0.55
Dut_Eng	conc	plicht	h	0.19	duty	v	0.45
Dut_Eng	conc	plug	v	0.40	plug	v	0.49
Dut_Eng	prop	pluizig	v	0.46	fluffy	h	0.41
Dut_Eng	conc	poeder	v	0.46	powder	v	0.24
Dut_Eng	conc	poëzie	a	0.11	poetry	a	0.53
Dut_Eng	conc	poging	v	0.03	attempt	v	0.41
Dut_Eng	conc	pond	v	0.41	pound	v	0.39
Dut_Eng	prop	prachtig	v	0.45	gorgeous	v	0.25

Dut_Eng	conc	president	v	0.43	president	v	0.46
Dut_Eng	conc	prijs	v	0.35	prize	v	0.33
Dut_Eng	prop	prikkelend	h	0.35	stinging	h	0.66
Dut_Eng	conc	primair	v	0.77	primary	v	0.39
Dut_Eng	conc	prins	v	0.17	prince	v	0.45
Dut_Eng	conc	probleem	v	0.07	problem	a	0.35
Dut_Eng	conc	productie	v	0.12	production	v	0.24
Dut_Eng	conc	promotie	v	0.28	promotion	a	0.41
Dut_Eng	conc	proportie	v	0.48	proportion	v	0.32
Dut_Eng	prop	pulserend	h	0.18	pulsing	h	0.36
Dut_Eng	prop	puntig	v	0.42	spiky	h	0.43
Dut_Eng	conc	pupil	v	0.56	pupil	v	0.46
Dut_Eng	prop	puur	v	0.14	sheer	v	0.57
Dut_Eng	conc	raad	a	0.44	council	v	0.45
Dut_Eng	conc	race	v	0.34	race	v	0.50
Dut_Eng	conc	raket	v	0.28	missile	v	0.41
Dut_Eng	conc	rand	v	0.44	edge	v	0.46
Dut_Eng	prop	raspend	v	0.17	raspy	a	0.39
Dut_Eng	prop	rauw	v	0.46	raucous	a	0.52
Dut_Eng	prop	recht	v	0.49	straight	v	0.53
Dut_Eng	prop	rechthoekig	v	0.54	rectangular	v	0.54
Dut_Eng	conc	reflectie	v	0.54	reflection	v	0.84
Dut_Eng	conc	rekening	v	0.48	account	a	0.42
Dut_Eng	conc	relatie	v	0.09	relation	v	0.39
Dut_Eng	conc	republiek	v	0.36	republic	v	0.63
Dut_Eng	conc	reus	v	0.33	giant	v	0.43
Dut_Eng	prop	reusachtig	v	0.68	huge	v	0.43
Dut_Eng	conc	richting	v	0.27	direction	v	0.43
Dut_Eng	conc	riem	v	0.57	belt	v	0.41
Dut_Eng	prop	rinkelend	a	0.52	jingling	a	0.52
Dut_Eng	prop	ritmisch	a	0.20	rhythmic	a	0.54
Dut_Eng	prop	ritselend	a	0.31	rustling	a	0.41
Dut_Eng	prop	robuust	v	0.31	sturdy	h	0.41

Dut_Eng	prop	roestig	v	0.43	rusty	v	0.34
Dut_Eng	conc	rok	h	0.41	skirt	v	0.49
Dut_Eng	conc	rol	v	0.08	role	v	0.48
Dut_Eng	prop	rommelend	v	0.22	rumbling	a	0.46
Dut_Eng	prop	rond	v	0.47	round	v	0.48
Dut_Eng	prop	rood	v	0.74	red	v	0.83
Dut_Eng	prop	roodachtig	v	0.74	reddish	v	0.86
Dut_Eng	prop	roosterend	v	0.32	roasting	v	0.17
Dut_Eng	prop	rot	v	0.47	rotten	v	0.28
Dut_Eng	conc	rots	h	0.35	rock	v	0.41
Dut_Eng	prop	rotsig	v	0.36	craggy	v	0.45
Dut_Eng	conc	route	v	0.45	route	v	0.51
Dut_Eng	prop	roze	v	0.65	pink	v	0.79
Dut_Eng	prop	rubberachtig	v	0.24	rubbery	h	0.25
Dut_Eng	prop	ruig	v	0.22	shaggy	v	0.49
Dut_Eng	prop	ruim	v	0.55	large	v	0.39
Dut_Eng	prop	ruisend	a	0.57	murmuring	a	0.63
Dut_Eng	prop	rustig	v	0.27	quiet	a	0.67
Dut_Eng	prop	ruw	h	0.32	rough	h	0.31
Dut_Eng	conc	ruzie	v	0.15	quarrel	a	0.50
Dut_Eng	prop	saai	v	0.15	drab	v	0.39
Dut_Eng	conc	schade	v	0.08	harm	v	0.33
Dut_Eng	prop	schattig	v	0.49	cute	v	0.41
Dut_Eng	conc	schatting	v	0.12	estimate	v	0.27
Dut_Eng	prop	scheef	v	0.46	crooked	v	0.48
Dut_Eng	prop	schel	a	0.48	shrill	a	0.69
Dut_Eng	conc	schelp	v	0.23	shell	v	0.31
Dut_Eng	prop	scherp	v	0.22	sharp	h	0.37
Dut_Eng	prop	schetterend	a	0.64	blaring	a	0.65
Dut_Eng	prop	schilferig	v	0.23	flaky	h	0.33
Dut_Eng	prop	schimmig	v	0.45	shadowy	v	0.82
Dut_Eng	prop	schitterend	v	0.41	dazzling	v	0.77
Dut_Eng	conc	schok	h	0.08	shock	v	0.35

Dut_Eng	conc	school	v	0.19	school	v	0.36
Dut_Eng	prop	schoon	v	0.31	clean	v	0.29
Dut_Eng	conc	schoonheid	v	0.24	beauty	v	0.33
Dut_Eng	prop	schor	a	0.59	husky	a	0.57
Dut_Eng	prop	schreeuwend	a	0.22	shrieking	a	0.59
Dut_Eng	prop	schreeuwerig	a	0.54	clamorous	h	0.32
Dut_Eng	prop	schriel	v	0.40	scrawny	v	0.52
Dut_Eng	conc	schrift	v	0.40	writing	v	0.60
Dut_Eng	prop	schuimig	v	0.42	foamy	v	0.42
Dut_Eng	conc	schuld	v	0.05	guilt	v	0.38
Dut_Eng	prop	schurend	h	0.10	abrasive	h	0.33
Dut_Eng	conc	seconde	a	0.11	second	v	0.45
Dut_Eng	conc	selectie	v	0.21	selection	v	0.18
Dut_Eng	conc	sergeant	v	0.36	sergeant	v	0.51
Dut_Eng	conc	sierlijkheid	v	0.30	grace	v	0.43
Dutch	conc	sirene	a	0.47	siren		
Dut_Eng	prop	sissend	a	0.37	hissing	a	0.64
Dut_Eng	conc	situatie	v	0.28	situation	v	0.30
Dut_Eng	prop	slap	h	0.42	floppy	v	0.46
Dut_Eng	conc	slavernij	v	0.19	slavery	v	0.41
Dut_Eng	prop	slijmerig	h	0.29	slimy	h	0.31
Dut_Eng	prop	slissend	a	0.46	gooey	h	0.29
Dut_Eng	prop	smakeloos	v	0.20	insipid	v	0.30
Dut_Eng	prop	smerig	v	0.31	filthy	v	0.47
Dut_Eng	conc	snee	v	0.43	cut	v	0.36
Dut_Eng	prop	snel	v	0.38	swift	v	0.50
Dut_Eng	conc	socialist	v	0.20	socialist	a	0.44
Dut_Eng	prop	soepel	v	0.33	smooth	h	0.36
Dut_Eng	prop	solide	v	0.34	solid	h	0.35
Dut_Eng	prop	somber	v	0.40	dim	v	0.75
Dut_Eng	prop	sonoor	h	0.41	sonorous	a	0.74
Dut_Eng	conc	soort	v	0.32	kind	v	0.33
Dut_Eng	conc	spanning	h	0.01	strain	h	0.33

Dut_Eng	conc	spier	h	0.29	muscle	h	0.40
Dut_Eng	prop	spinnend	a	0.36	purring	a	0.59
Dut_Eng	prop	spits	v	0.49	tapering	v	0.43
Dutch	conc	spons	h	0.42	sponge		
Dut_Eng	conc	spoor	v	0.34	track	v	0.49
Dut_Eng	prop	sprankelend	v	0.40	sparkly	v	0.60
Dut_Eng	conc	stad	v	0.29	city	v	0.26
Dut_Eng	prop	stampend	a	0.18	thumping	a	0.40
Dut_Eng	prop	steil	v	0.55	steep	v	0.57
Dut_Eng	prop	stekelig	h	0.29	prickly	h	0.40
Dut_Eng	conc	stem	a	0.64	vote	v	0.44
Dut_Eng	prop	sterk	v	0.30	strong	v	0.13
Dut_Eng	conc	sterkte	h	0.11	strength	v	0.36
Dut_Eng	conc	sterven	v	0.31	die	v	0.34
Dut_Eng	prop	stevig	v	0.23	tough	h	0.36
Dutch	conc	stijfheid	h	0.32	stiffness		
Dut_Eng	conc	stijgen	v	0.23	rise	v	0.43
Dut_Eng	prop	stil	a	0.44	silent	a	0.70
Dut_Eng	prop	stilstaand	v	0.41	stagnant	v	0.31
Dut_Eng	conc	stoel	h	0.31	chair	v	0.46
Dut_Eng	conc	stof	v	0.30	dust	v	0.47
Dut_Eng	prop	stoffig	v	0.40	dusty	v	0.46
Dut_Eng	prop	stom	v	0.25	mute	a	0.75
Dut_Eng	prop	stomp	v	0.33	blunt	h	0.43
Dut_Eng	conc	stop	v	0.15	stop	v	0.34
Dut_Eng	prop	stormachtig	a	0.13	stormy	v	0.41
Dut_Eng	prop	strak	v	0.36	tight	h	0.47
Dut_Eng	prop	stralend	v	0.52	radiant	v	0.50
Dutch	conc	streling	h	0.59	caress		
Dut_Eng	conc	student	v	0.18	student	v	0.40
Dut_Eng	conc	stuk	v	0.29	piece	v	0.28
Dut_Eng	conc	substituut	v	0.35	substitute	v	0.30
Dut_Eng	conc	succes	v	0.31	success	v	0.38

Dut_Eng	prop	swingend	v	0.18	swinging	v	0.49
Dut_Eng	conc	symbool	v	0.69	symbol	v	0.49
Dut_Eng	prop	taai	h	0.30	wiry	v	0.38
Dut_Eng	conc	talent	v	0.24	talent	v	0.27
Dut_Eng	conc	tante	v	0.30	aunt	v	0.32
Dut_Eng	conc	team	v	0.20	team	v	0.38
Dut_Eng	prop	teerachtig	v	0.37	tarry	v	0.43
Dut_Eng	conc	tegenstand	v	0.31	opposition	v	0.44
Dut_Eng	conc	tekst	v	0.36	text	v	0.60
Dut_Eng	conc	telefoon	a	0.08	phone	a	0.37
Dut_Eng	conc	telefoontje	a	0.18	call	a	0.56
Dut_Eng	prop	tenger	v	0.42	petite	v	0.60
Dut_Eng	conc	terug	v	0.33	back	v	0.44
Dut_Eng	conc	tevredenheid	v	0.27	satisfaction	v	0.15
Dut_Eng	conc	thema	v	0.21	theme	v	0.35
Dut_Eng	conc	theorie	v	0.31	theory	a	0.38
Dut_Eng	prop	tijlpend	a	0.64	beeping	a	0.75
Dut_Eng	conc	tikje	h	0.24	pat	h	0.44
Dut_Eng	prop	tinkelend	a	0.17	tinkling	a	0.54
Dut_Eng	prop	tintelend	h	0.54	tingly	h	0.57
Dut_Eng	conc	titel	v	0.49	title	v	0.57
Dut_Eng	conc	toelating	v	0.24	admission	v	0.52
Dut_Eng	conc	toename	v	0.37	increase	v	0.25
Dut_Eng	conc	toestemming	a	0.30	permission	a	0.42
Dut_Eng	conc	toevoeging	v	0.39	addition	v	0.41
Dut_Eng	conc	tractor	v	0.28	tractor	v	0.36
Dut_Eng	prop	transparant	v	0.61	transparent	v	0.78
Dutch	conc	trekker	v	0.20	trigger		
Dut_Eng	conc	trend	v	0.41	trend	v	0.41
Dut_Eng	prop	troebel	v	0.72	murky	v	0.51
Dut_Eng	conc	trots	v	0.28	pride	v	0.45
Dut_Eng	conc	type	v	0.27	type	v	0.15
Dut_Eng	conc	uitrusting	v	0.44	equipment	v	0.37

Dut_Eng	conc	uitvinding	v	0.07	invention	v	0.26
Dut_Eng	conc	uitwisseling	v	0.22	exchange	v	0.33
Dut_Eng	conc	uitzicht	v	0.66	view	v	0.65
Dut_Eng	conc	uur	v	0.42	hour	v	0.59
Dut_Eng	prop	vaag	v	0.30	fuzzy	v	0.43
Dut_Eng	conc	vacuüm	v	0.11	vacuum	v	0.35
Dut_Eng	prop	vallend	v	0.23	falling	v	0.46
Dut_Eng	conc	veiligheid	v	0.32	safety	v	0.40
Dut_Eng	conc	vel	h	0.36	fell	v	0.39
Dut_Eng	conc	venster	v	0.66	window	v	0.49
Dut_Eng	conc	verbeelding	v	0.37	imagination	v	0.23
Dut_Eng	conc	verf	v	0.38	paint	v	0.34
Dut_Eng	conc	vergadering	a	0.39	meeting	v	0.41
Dut_Eng	conc	verhoor	a	0.43	trial	v	0.39
Dut_Eng	conc	verhulling	v	0.37	cope	v	0.42
Dut_Eng	conc	verklaring	a	0.29	statement	v	0.41
Dut_Eng	prop	verkoold	v	0.33	charred	v	0.31
Dut_Eng	conc	verleden	a	0.15	past	a	0.26
Dut_Eng	conc	verloving	v	0.15	engagement	v	0.41
Dut_Eng	conc	vermindering	v	0.26	reduction	v	0.33
Dut_Eng	prop	verouderd	v	0.45	wizened	v	0.42
Dut_Eng	conc	verpleger	v	0.30	orderly	v	0.50
Dut_Eng	conc	verscheidenheid	v	0.21	variety	v	0.08
Dut_Eng	prop	verstild	a	0.16	hushed	a	0.69
Dut_Eng	prop	vertakkend	v	0.38	branching	v	0.51
Dut_Eng	conc	vertrouwen	v	0.02	trust	v	0.39
Dut_Eng	conc	verzameling	v	0.31	collection	v	0.41
Dut_Eng	prop	vetachtig	v	0.28	greasy	h	0.27
Dut_Eng	prop	vierkant	v	0.47	square	v	0.52
Dut_Eng	prop	vies	v	0.56	grubby	v	0.48
Dutch	conc	vijl	v	0.13	file		
Dut_Eng	conc	vilt	h	0.40	felt	h	0.47
Dut_Eng	conc	vinden	v	0.06	find	v	0.30

Dut_Eng	prop	vlak	v	0.44	flat	v	0.44
Dut_Eng	conc	vlees	v	0.32	quick	v	0.47
Dut_Eng	prop	vlekkeloos	v	0.72	spotless	v	0.63
Dut_Eng	prop	vlezig	v	0.36	fleshy	v	0.38
Dut_Eng	conc	vliegtuig	v	0.14	plane	v	0.32
Dut_Eng	conc	vloeistof	h	0.27	liquid	v	0.09
Dut_Eng	prop	vochtig	v	0.25	humid	h	0.42
Dut_Eng	conc	voet	h	0.26	foot	v	0.35
Dut_Eng	conc	voeten	v	0.22	feet	v	0.32
Dut_Eng	conc	volume	a	0.48	volume	a	0.55
Dut_Eng	prop	volumineus	v	0.14	bulky	v	0.47
Dut_Eng	conc	vondst	v	0.23	finish	v	0.33
Dut_Eng	conc	voorbereiding	v	0.23	preparation	v	0.23
Dut_Eng	conc	voordeel	v	0.17	advantage	v	0.45
Dut_Eng	conc	voornemen	a	0.31	intention	v	0.46
Dut_Eng	conc	voorziening	a	0.23	provision	v	0.33
Dut_Eng	conc	vorm	v	0.46	form	v	0.40
Dut_Eng	conc	vraag	a	0.33	question	a	0.63
Dut_Eng	prop	vreemd	v	0.16	strange	v	0.10
Dut_Eng	conc	vreemdeling	v	0.35	stranger	v	0.40
Dut_Eng	conc	vriend	a	0.04	peer	v	0.31
Dut_Eng	conc	vrijheid	v	0.20	liberty	v	0.43
Dut_Eng	conc	vrijwilliger	v	0.21	voluntary	v	0.56
Dut_Eng	prop	vuil	v	0.27	dirty	v	0.34
Dut_Eng	conc	waarde	v	0.13	value	v	0.31
Dut_Eng	conc	wacht	v	0.40	hold	h	0.52
Dut_Eng	conc	wachten	v	0.28	wait	v	0.43
Dut_Eng	conc	wantrouwen	a	0.19	suspicion	v	0.40
Dut_Eng	prop	warm	h	0.39	warm	h	0.35
Dut_Eng	prop	wasachtig	h	0.39	waxy	h	0.31
Dut_Eng	prop	wazig	v	0.65	hazy	v	0.74
Dut_Eng	prop	weelderig	v	0.65	lush	v	0.23
Dut_Eng	conc	weer	v	0.29	weather	v	0.35

Dut_Eng	prop	weerkaatsend	v	0.32	reverberating	a	0.43
Dut_Eng	conc	weerstand	v	0.16	resistance	v	0.38
Dut_Eng	prop	weinig	v	0.21	little	v	0.56
Dut_Eng	conc	welkom	a	0.19	welcome	a	0.30
Dut_Eng	conc	welzijn	h	0.16	welfare	v	0.35
Dut_Eng	conc	werk	v	0.21	work	v	0.29
Dut_Eng	conc	westen	v	0.39	west	v	0.61
Dut_Eng	conc	wiel	v	0.27	wheel	v	0.45
Dut_Eng	prop	wild	v	0.22	gamy	h	0.23
Dutch	conc	wind	h	0.19	wind		
Dut_Eng	prop	wit	v	0.80	white	v	0.87
Dut_Eng	prop	wollig	v	0.38	woolly	h	0.47
Dut_Eng	conc	worp	v	0.24	throw	v	0.47
Dut_Eng	conc	wrok	a	0.09	spite	a	0.52
Dut_Eng	conc	zaad	h	0.51	seed	v	0.31
Dut_Eng	prop	zacht	h	0.24	soft	h	0.34
Dut_Eng	conc	zaken	v	0.24	business	v	0.48
Dut_Eng	prop	zanderig	v	0.13	gritty	h	0.35
Dut_Eng	prop	zangerig	a	0.67	lilting	v	0.31
Dut_Eng	conc	zee	v	0.10	sea	v	0.10
Dut_Eng	conc	zeep	h	0.34	soap	h	0.27
Dut_Eng	prop	zeepachtig	v	0.38	soapy	v	0.24
Dut_Eng	prop	zeer	h	0.36	sore	h	0.48
Dut_Eng	conc	zeer	h	0.16	hurt	v	0.32
Dutch	conc	zenuw	h	0.34	nerve		
Dut_Eng	prop	zeurend	a	0.27	whining	a	0.62
Dut_Eng	conc	zicht	v	0.47	sight	v	0.80
Dut_Eng	conc	ziekte	v	0.09	disease	v	0.23
Dutch	conc	zijde	h	0.45	silk		
Dut_Eng	prop	zijdeachtig	h	0.36	silky	h	0.53
Dut_Eng	prop	zilveren	v	0.58	silver	v	0.77
Dut_Eng	prop	zoemend	a	0.60	buzzing	a	0.55
Dut_Eng	prop	zonnig	v	0.33	sunny	v	0.74

Dut_Eng	prop	zuchtend	a	0.45	moaning	a	0.57
Dut_Eng	conc	zus	v	0.18	sister	a	0.29
Dut_Eng	prop	zwaar	h	0.14	heavy	h	0.37
Dut_Eng	prop	zwak	v	0.28	weak	v	0.18
Dut_Eng	prop	zwart	v	0.51	black	v	0.89
Dut_Eng	prop	zwoel	v	0.25	muggy	v	0.34

Appendix 3 R code for the analysis, with the results printed. Rendered automatically from the original script

```
# READ-ME
# The 'all.csv' file, created outside of R, in Excel, compiles all individual ratings.
# Dutch and English data are described in separate columns. All analyses separate for
# properties and concepts, except for a translation check.
# Stat tests (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).
# The code is extensively annotated, but some clarifications are in order. Subsetting is
# done throughout the code, and is essential due to the different norms (see 'normed'
# column: English, Dutch, or both). Subsetting is often done on the basis of variables
# that are unique to either norms, especially, 'Exclusivity' and 'exc_eng'.
# At first, the code must be run right from the top, as different objects bear the
# same name. In its entirety, it takes ~20 mins. Note the annotations for theoretical
# matters. Long variables are never presented entirely, but rather in sections or via
# summaries. Yet, the reader is invited to edit and present them entirely.
# Written on R version 3.2.2 (2015-08-14). This script markdown presents each code chunk
# followed by the results.


# INDEX
# Libraries: please ensure every library loads, and otherwise install it via
#   install.packages("")
# Preprocessing
# Translation-dependent results
# Critical results
#     Modality
#     Sound-symbolism


# _____________________ ---  START  --- _____________________

# Set your working directory here, to yield figures output:
# setwd('C:/.../.../...')

install.packages("gdata")

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'gdata' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages("GPArotation")
```

```
## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'GPArotation' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages("psych")

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'psych' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages("ggplot2")

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages("car")

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'car' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages("Rmisc")

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'Rmisc' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages("corpcor")

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)
```

```
## package 'corpcor' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages
```

```r
install.packages('contrast')
```

```
## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'contrast' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages
```

```r
install.packages('doBy')
```

```
## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'doBy' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages
```

```r
install.packages('ltm')
```

```
## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'ltm' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages
```

```r
install.packages('MASS')
```

```
## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'MASS' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages
```

```r
install.packages('QuantPsyc')
```

```
## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'QuantPsyc' successfully unpacked and MD5 sums checked
##
```

```
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('qpcR')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'qpcR' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('corpcor')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'corpcor' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('lattice')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'lattice' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('car')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'car' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('pastecs')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'pastecs' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages
```

```r
install.packages('scales')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'scales' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('reshape')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'reshape' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('arules')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'arules' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('plyr')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'plyr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('RColorBrewer')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'RColorBrewer' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('dplyr')
```

```
## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'dplyr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('gdata')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'gdata' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('gtools')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'gtools' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('Hmisc')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'Hmisc' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

install.packages('png')

## Installing package into 'C:/Users/Pablo/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'png' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Pablo\AppData\Local\Temp\Rtmp4so1up\downloaded_packages

library(ltm)

## Loading required package: MASS

## Loading required package: msm
```

```
## Loading required package: polycor

## Loading required package: mvtnorm

## Loading required package: sfsmisc

library(lattice)
library(psych)

##
## Attaching package: 'psych'

## The following object is masked from 'package:ltm':
##
##     factor.scores

## The following object is masked from 'package:polycor':
##
##     polyserial

library(car)

##
## Attaching package: 'car'

## The following object is masked from 'package:psych':
##
##     logit

library(doBy)
library(contrast)

## Loading required package: rms

## Loading required package: Hmisc

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##     %+%, alpha

##
## Attaching package: 'Hmisc'
```

```
## The following object is masked from 'package:psych':
##
##     describe

## The following object is masked from 'package:sfsmisc':
##
##     errbar

## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units

## Loading required package: SparseM

##
## Attaching package: 'SparseM'

## The following object is masked from 'package:base':
##
##     backsolve

##
## Attaching package: 'rms'

## The following object is masked from 'package:car':
##
##     vif

library(pastecs)

## Loading required package: boot

##
## Attaching package: 'boot'

## The following object is masked from 'package:survival':
##
##     aml

## The following object is masked from 'package:car':
##
##     logit

## The following object is masked from 'package:psych':
##
##     logit

## The following object is masked from 'package:lattice':
##
##     melanoma
```

```
## The following object is masked from 'package:msm':
##
##     cav

##
## Attaching package: 'pastecs'

## The following object is masked from 'package:rms':
##
##     specs

## The following object is masked from 'package:sfsmisc':
##
##     last

library(scales)

##
## Attaching package: 'scales'

## The following objects are masked from 'package:psych':
##
##     alpha, rescale

library(ggplot2)
library(psych)
library(reshape)
library(arules)

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following object is masked from 'package:reshape':
##
##     expand

##
## Attaching package: 'arules'

## The following object is masked from 'package:car':
##
##     recode

## The following objects are masked from 'package:base':
##
##     abbreviate, write

library(plyr)

##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:reshape':
##
##     rename, round_any

## The following objects are masked from 'package:Hmisc':
##
##     is.discrete, summarize

library(RColorBrewer)
library(Rmisc)
library(corpcor)
library(GPArotation)
library(gdata)

## gdata: Unable to locate valid perl interpreter
## gdata:
## gdata: read.xls() will be unable to read Excel XLS and XLSX files
## gdata: unless the 'perl=' argument is used to specify the location
## gdata: of a valid perl intrpreter.
## gdata:
## gdata: (To avoid display of this message in the future, please
## gdata: ensure perl is installed and available on the executable
## gdata: search path.)

## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLX' (Excel 97-2004) files.

##

## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLSX' (Excel 2007+) files.

##

## gdata: Run the function 'installXLSXsupport()'
## gdata: to automatically download and install the perl
## gdata: libaries needed to support Excel XLS and XLSX formats.

##
## Attaching package: 'gdata'

## The following objects are masked from 'package:pastecs':
##
##     first, last

## The following object is masked from 'package:Hmisc':
##
##     combine

## The following object is masked from 'package:sfsmisc':
##
##     last
```

```
## The following object is masked from 'package:stats':
##
##     nobs

## The following object is masked from 'package:utils':
##
##     object.size

## The following object is masked from 'package:base':
##
##     startsWith

library(QuantPsyc)

##
## Attaching package: 'QuantPsyc'

## The following object is masked from 'package:Matrix':
##
##     norm

## The following object is masked from 'package:SparseM':
##
##     norm

## The following object is masked from 'package:base':
##
##     norm

library(MASS)
library(qpcR)

## Loading required package: minpack.lm

## Loading required package: rgl

## Loading required package: robustbase

##
## Attaching package: 'robustbase'

## The following object is masked from 'package:boot':
##
##     salinity

## The following object is masked from 'package:survival':
##
##     heart

## The following object is masked from 'package:psych':
##
##     cushny
```

```r
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:gdata':
##
##     combine, first, last

## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following objects are masked from 'package:arules':
##
##     intersect, recode, setdiff, setequal, union

## The following object is masked from 'package:reshape':
##
##     rename

## The following objects are masked from 'package:pastecs':
##
##     first, last

## The following objects are masked from 'package:Hmisc':
##
##     combine, src, summarize

## The following object is masked from 'package:car':
##
##     recode

## The following object is masked from 'package:sfsmisc':
##
##     last

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(gtools)
```

```
##
## Attaching package: 'gtools'

## The following objects are masked from 'package:boot':
##
##     inv.logit, logit

## The following object is masked from 'package:car':
##
##     logit

## The following object is masked from 'package:psych':
##
##     logit

library(Hmisc)
library(png)


# Calculate average percentange of unresponded items, i.e., unknown. Since there are
# three ratings per word, and indeed the three were left blank whereever participants
# ignored some word, the calculation includes a division by 3 (besides overall mean,
# see specific percentage per file).

file1 <- read.csv('file1_gral.csv')
file2 <- read.csv('file2_gral.csv')
file3 <- read.csv('file3_gral.csv')
file4 <- read.csv('file4_gral.csv')
file5 <- read.csv('file5_gral.csv')
file6 <- read.csv('file6_gral.csv')

(((100 * (sum(is.na(file1)))) / (sum(!is.na(file1[,-1])) + sum(is.na(file1))) /3) +
# 0.29

((100 * (sum(is.na(file2)))) / (sum(!is.na(file2[,-1])) + sum(is.na(file2))) /3) +
# 1.42

((100 * (sum(is.na(file3)))) / (sum(!is.na(file3[,-1])) + sum(is.na(file3))) /3) +
# 0.41

# N.B. First participant is ignored because she completed only the first half of the
# survey.
((100 * (sum(is.na(file4[,-c(1:4)])))) / (sum(!is.na(file4[,-c(1:4)])) +
sum(is.na(file4[,-c(1:4)]))) /3) +
# 2.85

((100 * (sum(is.na(file5)))) / (sum(!is.na(file5[,-1])) + sum(is.na(file5))) /3) +
# 1.38

((100 * (sum(is.na(file6)))) / (sum(!is.na(file6[,-1])) + sum(is.na(file6))) /3)) /6
```

```
## [1] 1.308878

# 1.50
# /6 = 1.31% = average unknown
#_________________________________________________________________________________

# Preprocessing:
# There were 9 files with different items (mostly unrepeated) for concepts and
# 10 files for properties. They were completed in different proportions, with an
# average of eight participants per file.

# RELIABILITY ANALYSIS: In putting together the ratings from each respondent, this
# analysis allows to calculate the fit among those. In other words, is the mean
# realistic or forced? Two measures are provided. First, interitem consistency
# provides the fit among items independently of raters. Second, interrater
# reliability measures the fit among raters, independently of items. A standard
# minimum for both is alpha = .70.

# Concepts
all <- read.csv('all.csv')
concs <- all[all$cat == 'conc',]

# There were
a_concs<-concs[, c('a1', 'a2', 'a3', 'a4', 'a5', 'a6', 'a7', 'a8', 'a9')]
psych::alpha(a_concs)

##
## Reliability analysis
## Call: psych::alpha(x = a_concs)
##
##   raw_alpha std.alpha G6(smc) average_r S/N   ase mean sd
##       0.74      0.74    0.75      0.24 2.9 0.019    2  1
##
##  lower alpha upper     95% confidence boundaries
## 0.7 0.74 0.78
##
##  Reliability if an item is dropped:
##    raw_alpha std.alpha G6(smc) average_r S/N alpha se
## a1      0.72      0.72    0.73      0.24 2.5    0.021
## a2      0.73      0.73    0.74      0.25 2.7    0.020
## a3      0.73      0.73    0.74      0.26 2.7    0.020
## a4      0.71      0.71    0.71      0.24 2.5    0.021
## a5      0.71      0.71    0.72      0.24 2.5    0.021
## a6      0.73      0.74    0.75      0.26 2.8    0.020
## a7      0.70      0.70    0.71      0.23 2.4    0.022
## a8      0.71      0.72    0.71      0.24 2.5    0.021
## a9      0.70      0.70    0.71      0.23 2.4    0.022
##
##  Item statistics
##      n raw.r std.r r.cor r.drop mean  sd
```

```
## a1 398  0.59  0.57  0.49   0.42  2.9 1.9
## a2 408  0.51  0.52  0.41   0.35  2.0 1.6
## a3 410  0.54  0.50  0.40   0.33  1.9 1.9
## a4 409  0.61  0.60  0.54   0.45  2.4 1.6
## a5 409  0.60  0.60  0.54   0.44  1.3 1.9
## a6 407  0.47  0.48  0.36   0.31  2.2 1.9
## a7 410  0.65  0.65  0.60   0.51  1.9 1.9
## a8 269  0.56  0.59  0.54   0.43  1.4 1.6
## a9 263  0.64  0.64  0.59   0.51  1.5 1.8
```

```r
h_concs<-concs[, c('h1', 'h2', 'h3', 'h4', 'h5', 'h6', 'h7', 'h8', 'h9')]
psych::alpha(h_concs)
```

```
##
## Reliability analysis
## Call: psych::alpha(x = h_concs)
##
##   raw_alpha std.alpha G6(smc) average_r S/N  ase mean sd
##       0.72      0.72    0.74      0.22 2.6 0.02    2  1
##
##  lower alpha upper     95% confidence boundaries
## 0.68 0.72 0.76
##
##  Reliability if an item is dropped:
##    raw_alpha std.alpha G6(smc) average_r S/N alpha se
## h1      0.70      0.70    0.72      0.23 2.3    0.022
## h2      0.72      0.72    0.72      0.24 2.5    0.021
## h3      0.71      0.70    0.71      0.23 2.4    0.021
## h4      0.71      0.72    0.72      0.24 2.5    0.021
## h5      0.68      0.68    0.70      0.21 2.1    0.024
## h6      0.70      0.70    0.71      0.23 2.3    0.022
## h7      0.67      0.67    0.68      0.20 2.0    0.025
## h8      0.69      0.69    0.71      0.22 2.2    0.023
## h9      0.69      0.69    0.69      0.22 2.2    0.022
##
##  Item statistics
##      n raw.r std.r r.cor r.drop mean  sd
## h1 397  0.56  0.54  0.44   0.38  3.4 1.9
## h2 408  0.43  0.46  0.36   0.27  1.8 1.7
## h3 410  0.59  0.52  0.44   0.36  2.4 2.2
## h4 409  0.46  0.46  0.36   0.30  2.5 1.6
## h5 409  0.62  0.63  0.56   0.48  1.2 1.9
## h6 407  0.53  0.54  0.45   0.38  1.8 1.7
## h7 410  0.69  0.68  0.65   0.54  1.5 1.9
## h8 269  0.61  0.60  0.52   0.44  1.7 1.7
## h9 263  0.57  0.58  0.53   0.41  1.1 1.6
```

```r
v_concs<-concs[, c('v1', 'v2', 'v3', 'v4', 'v5', 'v6', 'v7', 'v8', 'v9')]
psych::alpha(v_concs)
```

```
##
## Reliability analysis
## Call: psych::alpha(x = v_concs)
##
##   raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd
##        0.7       0.7    0.72      0.21 2.4 0.022  3.1 0.95
##
##  lower alpha upper     95% confidence boundaries
## 0.66 0.7 0.75
##
##  Reliability if an item is dropped:
##    raw_alpha std.alpha G6(smc) average_r S/N alpha se
## v1      0.69      0.70    0.71      0.22 2.3    0.023
## v2      0.70      0.70    0.70      0.23 2.4    0.022
## v3      0.66      0.66    0.67      0.19 1.9    0.025
## v4      0.68      0.67    0.68      0.21 2.1    0.024
## v5      0.67      0.67    0.69      0.20 2.0    0.024
## v6      0.70      0.70    0.71      0.22 2.3    0.022
## v7      0.66      0.65    0.65      0.19 1.9    0.025
## v8      0.69      0.69    0.70      0.22 2.2    0.023
## v9      0.65      0.65    0.66      0.19 1.8    0.026
##
##  Item statistics
##      n raw.r std.r r.cor r.drop mean  sd
## v1 398  0.45  0.46  0.34   0.28  4.1 1.4
## v2 408  0.40  0.43  0.32   0.25  2.5 1.6
## v3 409  0.63  0.62  0.57   0.47  3.3 1.9
## v4 409  0.55  0.56  0.49   0.39  3.2 1.5
## v5 409  0.64  0.58  0.49   0.42  2.4 2.2
## v6 407  0.47  0.45  0.32   0.27  3.6 1.7
## v7 410  0.64  0.64  0.62   0.49  3.1 1.7
## v8 269  0.47  0.48  0.38   0.31  2.5 1.7
## v9 263  0.64  0.66  0.62   0.53  3.3 1.6

# RESULTS good. Lower than L&C, but they had more participants.
# Interitem consistency
# a: .74
# h: .72
# v: .70

# Interrater reliability
# a: .75
# h: .74
# v: .72



# Properties
props <- all[all$cat == 'prop',]
```

```r
a_props<-props[, c('a1', 'a2', 'a3', 'a4', 'a5', 'a6', 'a7', 'a8', 'a9', 'a10')]
psych::alpha(a_props)
```

```
## Warning in psych::alpha(a_props): Some items were negatively correlated
## with the total scale and probably should be reversed. To do this, run the
## function again with the 'check.keys=TRUE' option

## Some items ( a10 ) were negatively correlated with the total scale and probably should
be reversed.  To do this, run the function again with the 'check.keys=TRUE' option

##
## Reliability analysis
## Call: psych::alpha(x = a_props)
##
##   raw_alpha std.alpha G6(smc) average_r S/N   ase mean  sd
##       0.78      0.78    0.89      0.27 3.6 0.016  1.7 1.3
##
##  lower alpha upper     95% confidence boundaries
## 0.75 0.78 0.81
##
##  Reliability if an item is dropped:
##     raw_alpha std.alpha G6(smc) average_r S/N alpha se
## a1       0.71      0.72    0.83      0.22 2.6    0.022
## a2       0.74      0.75    0.87      0.25 2.9    0.019
## a3       0.77      0.78    0.87      0.28 3.5    0.017
## a4       0.72      0.73    0.85      0.23 2.6    0.021
## a5       0.79      0.79    0.87      0.30 3.8    0.015
## a6       0.72      0.72    0.85      0.22 2.6    0.021
## a7       0.73      0.74    0.86      0.24 2.8    0.020
## a8       0.73      0.73    0.86      0.23 2.7    0.020
## a9       0.80      0.81    0.90      0.33 4.4    0.016
## a10      0.85      0.84    0.90      0.37 5.4    0.011
##
##  Item statistics
##       n  raw.r  std.r r.cor r.drop mean  sd
## a1  256  0.863  0.840  0.86  0.805 1.32 2.0
## a2  305  0.755  0.713  0.67  0.638 1.61 1.7
## a3  319  0.490  0.501  0.45  0.345 2.51 1.9
## a4  320  0.844  0.829  0.83  0.760 1.35 1.9
## a5  321  0.440  0.399  0.35  0.220 2.20 2.1
## a6  214  0.882  0.855  0.86  0.800 1.36 1.7
## a7  213  0.808  0.771  0.76  0.669 1.42 2.0
## a8  217  0.824  0.803  0.79  0.727 2.00 1.7
## a9  109  0.225  0.205  0.10  0.035 0.91 1.1
## a10 109 -0.028 -0.087 -0.19 -0.258 2.23 2.1
##
## Non missing response frequency for each item
##        0 0.5    1    2    3    4    5 miss
## a1  0.64   0 0.05 0.04 0.05 0.05 0.16 0.25
## a2  0.42   0 0.07 0.21 0.15 0.07 0.08 0.11
```

```
## a3  0.25   0 0.08 0.15 0.14 0.18 0.20 0.07
## a4  0.54   0 0.12 0.09 0.06 0.03 0.15 0.07
## a5  0.43   0 0.02 0.08 0.12 0.11 0.25 0.06
## a6  0.49   0 0.17 0.08 0.10 0.08 0.08 0.38
## a7  0.64   0 0.03 0.02 0.07 0.06 0.18 0.38
## a8  0.14   0 0.45 0.09 0.08 0.08 0.16 0.37
## a9  0.53   0 0.14 0.25 0.06 0.01 0.01 0.68
## a10 0.42   0 0.04 0.03 0.13 0.17 0.21 0.68
```

```
h_props<-props[, c('h1', 'h2', 'h3', 'h4', 'h5', 'h6', 'h7', 'h8', 'h9', 'h10')]
psych::alpha(h_props)
```

```
## Warning in psych::alpha(h_props): Some items were negatively correlated
## with the total scale and probably should be reversed. To do this, run the
## function again with the 'check.keys=TRUE' option
```

```
## Some items ( h10 ) were negatively correlated with the total scale and probably should
be reversed.  To do this, run the function again with the 'check.keys=TRUE' option
```

```
##
## Reliability analysis
## Call: psych::alpha(x = h_props)
##
##   raw_alpha std.alpha G6(smc) average_r S/N   ase mean  sd
##        0.7      0.72    0.83       0.2 2.5 0.022    2 1.1
##
##  lower alpha upper     95% confidence boundaries
## 0.65 0.7 0.74
##
##  Reliability if an item is dropped:
##     raw_alpha std.alpha G6(smc) average_r S/N alpha se
## h1       0.60      0.65    0.76      0.17 1.8    0.029
## h2       0.64      0.67    0.80      0.19 2.0    0.026
## h3       0.70      0.72    0.82      0.22 2.6    0.022
## h4       0.66      0.68    0.80      0.19 2.2    0.025
## h5       0.69      0.71    0.81      0.21 2.5    0.023
## h6       0.61      0.63    0.77      0.16 1.7    0.029
## h7       0.60      0.64    0.79      0.16 1.8    0.031
## h8       0.63      0.65    0.78      0.17 1.9    0.027
## h9       0.70      0.73    0.83      0.23 2.7    0.022
## h10      0.81      0.81    0.87      0.32 4.2    0.014
##
##  Item statistics
##       n raw.r std.r r.cor r.drop mean  sd
## h1  256  0.82  0.77  0.79   0.69 2.28 2.0
## h2  306  0.67  0.65  0.62   0.54 1.62 1.7
## h3  319  0.36  0.38  0.30   0.21 2.63 1.6
## h4  320  0.61  0.60  0.54   0.44 1.48 1.9
## h5  321  0.50  0.45  0.40   0.27 2.10 2.1
## h6  214  0.79  0.82  0.83   0.73 1.78 1.7
## h7  213  0.82  0.80  0.80   0.71 2.21 2.1
```

```
## h8   217  0.71  0.75  0.75   0.62 2.54 1.6
## h9   109  0.37  0.32  0.22   0.13 0.96 1.0
## h10 109 -0.15 -0.24 -0.39  -0.40 1.72 2.2
##
## Non missing response frequency for each item
##         0    1 1.5    2    3 3.5    4    5 miss
## h1  0.36 0.05   0 0.09 0.19   0 0.11 0.21 0.25
## h2  0.42 0.09   0 0.16 0.16   0 0.11 0.06 0.11
## h3  0.15 0.11   0 0.20 0.21   0 0.18 0.16 0.07
## h4  0.52 0.09   0 0.09 0.10   0 0.05 0.14 0.07
## h5  0.41 0.05   0 0.11 0.11   0 0.08 0.23 0.06
## h6  0.38 0.08   0 0.12 0.27   0 0.08 0.07 0.38
## h7  0.43 0.03   0 0.05 0.12   0 0.17 0.21 0.38
## h8  0.08 0.29   0 0.11 0.17   0 0.20 0.15 0.37
## h9  0.36 0.45   0 0.10 0.07   0 0.00 0.02 0.68
## h10 0.61 0.00   0 0.00 0.06   0 0.09 0.24 0.68
```

```r
v_props<-props[, c('v1', 'v2', 'v3', 'v4', 'v5', 'v6', 'v7', 'v8', 'v9', 'v10')]
psych::alpha(v_props)
```

```
##
## Reliability analysis
## Call: psych::alpha(x = v_props)
##
##   raw_alpha std.alpha G6(smc) average_r S/N   ase mean  sd
##       0.85      0.85    0.87      0.36 5.5 0.012  3.2 1.2
##
##  lower alpha upper     95% confidence boundaries
## 0.82 0.85 0.87
##
##  Reliability if an item is dropped:
##     raw_alpha std.alpha G6(smc) average_r S/N alpha se
## v1       0.83      0.83    0.86      0.34 4.7    0.013
## v2       0.84      0.83    0.86      0.36 5.0    0.013
## v3       0.84      0.84    0.87      0.37 5.2    0.013
## v4       0.82      0.82    0.85      0.34 4.5    0.014
## v5       0.85      0.85    0.87      0.38 5.5    0.012
## v6       0.81      0.81    0.84      0.32 4.3    0.015
## v7       0.82      0.82    0.85      0.33 4.4    0.015
## v8       0.82      0.82    0.84      0.33 4.5    0.014
## v9       0.86      0.87    0.89      0.42 6.4    0.011
## v10      0.84      0.84    0.85      0.37 5.3    0.012
##
##  Item statistics
##       n raw.r std.r r.cor r.drop mean  sd
## v1  256  0.82  0.71  0.67   0.61  3.6 1.7
## v2  305  0.72  0.63  0.59   0.52  2.9 1.5
## v3  319  0.59  0.59  0.52   0.47  3.5 1.4
## v4  320  0.76  0.76  0.73   0.68  3.1 1.9
## v5  321  0.53  0.52  0.43   0.39  3.1 1.7
```

```
## v6  214  0.83  0.84  0.84   0.78  3.1 1.5
## v7  213  0.80  0.79  0.77   0.72  3.6 1.8
## v8  217  0.73  0.78  0.78   0.72  3.7 1.4
## v9  109  0.30  0.31  0.18   0.16  0.9 1.3
## v10 109  0.67  0.57  0.54   0.45  3.8 1.9
##
## Non missing response frequency for each item
##         0    1    2    3 3.5    4 4.333333333 4.5    5 miss
## v1  0.14 0.02 0.05 0.15   0 0.20           0   0 0.44 0.25
## v2  0.09 0.08 0.22 0.22   0 0.27           0   0 0.13 0.11
## v3  0.04 0.06 0.15 0.17   0 0.24           0   0 0.34 0.07
## v4  0.18 0.05 0.11 0.18   0 0.11           0   0 0.36 0.07
## v5  0.12 0.06 0.14 0.20   0 0.17           0   0 0.30 0.06
## v6  0.05 0.14 0.10 0.29   0 0.21           0   0 0.22 0.38
## v7  0.15 0.02 0.04 0.14   0 0.15           0   0 0.50 0.38
## v8  0.00 0.12 0.07 0.12   0 0.35           0   0 0.33 0.37
## v9  0.58 0.17 0.07 0.15   0 0.00           0   0 0.03 0.68
## v10 0.17 0.01 0.02 0.06   0 0.14           0   0 0.61 0.68

# RESULTS: good. Lower than L&C, but they did have a few more participants.
# Interitem consistency
# a: .78
# h: .70
# v: .85

# Interrater reliability
# a: .89
# h: .83
# v: .87
# ______________________________________________________________________________


# TRANSLATION-DEPENDENT RESULTS

# Read in the general norms file, 'all' (all seemed right, seeing as we have Dutch
# and English properties and concepts)

all <- read.csv('all.csv')

# PROPERTIES
props <- all[all$cat=='prop',]
nrow(props)  # 366 Dutch + a few from Lynott&Connell for comparisons

## [1] 343

# CONCEPTS
concs <- all[all$cat=='conc',]
nrow(concs)  # 411 Dutch + a few from Lynott&Connell for comparison

## [1] 416
```

```r
# Correlations

# PROPERTIES
# Modalities
rcor.test(props[, c('Auditory', 'Aud_eng')], use = 'complete.obs')

##
##          Auditory Aud_eng
## Auditory  *****    0.795
## Aud_eng  <0.001    *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

rcor.test(props[, c('Haptic', 'Hap_eng')], use = 'complete.obs')

##
##         Haptic Hap_eng
## Haptic   *****  0.690
## Hap_eng <0.001  *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

rcor.test(props[, c('Visual', 'Vis_eng')], use = 'complete.obs')

##
##         Visual Vis_eng
## Visual   *****  0.711
## Vis_eng <0.001  *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

# Significant, large correlations ranging from .69 to .80

# Exclusivity
rcor.test(props[, c('Exclusivity', 'exc_eng')], use = 'complete.obs')

##
##             Exclusivity exc_eng
## Exclusivity  *****       0.475
## exc_eng     <0.001       *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

# Medium-sized corr Eng-Dutch


# CONCEPTS
```

```r
# Modalities
rcor.test(concs[, c('Auditory', 'Aud_eng')], use = 'complete.obs')

##
##          Auditory Aud_eng
## Auditory  *****    0.683
## Aud_eng  <0.001    *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

rcor.test(concs[, c('Haptic', 'Hap_eng')], use = 'complete.obs')

##
##          Haptic Hap_eng
## Haptic   *****  0.624
## Hap_eng <0.001  *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

rcor.test(concs[, c('Visual', 'Vis_eng')], use = 'complete.obs')

##
##          Visual Vis_eng
## Visual   *****  0.659
## Vis_eng <0.001  *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

# Significant, large correlations ranging from .63 to .69

# Exclusivity
rcor.test(concs[, c('Exclusivity', 'exc_eng')], use = 'complete.obs')

##
##             Exclusivity exc_eng
## Exclusivity  *****       0.428
## exc_eng     <0.001       *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

# Medium-sized corr Eng-Dutch


# Descriptives: M, SD, SE...
# English
psych::describe(props$Aud_eng)
```

```
##    vars   n mean   sd median trimmed mad min max range skew kurtosis   se
## X1    1 343 1.73 1.67   1.05    1.54 1.2   0   5     5 0.86    -0.77 0.09

psych::describe(props$Hap_eng)

##    vars   n mean   sd median trimmed  mad min  max range  skew kurtosis
## X1    1 343 2.41 1.62   2.52    2.41 2.26   0 4.95  4.95 -0.07    -1.48
##      se
## X1 0.09

psych::describe(props$Vis_eng)

##    vars   n mean   sd median trimmed  mad  min max range  skew kurtosis
## X1    1 343  3.8 1.06   4.17    3.94 0.81 0.52   5  4.48 -1.07     0.16
##      se
## X1 0.06

psych::describe(concs$Aud_eng)

##    vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 392 2.16 1.09   2.06    2.12 1.13   0   5     5 0.34    -0.49 0.06

psych::describe(concs$Hap_eng)

##    vars   n mean   sd median trimmed  mad min  max range skew kurtosis
## X1    1 392 1.86 1.13   1.65    1.78 1.16   0 4.76  4.76 0.57    -0.63
##      se
## X1 0.06

psych::describe(concs$Vis_eng)

##    vars   n mean  sd median trimmed  mad  min max range skew kurtosis   se
## X1    1 392 3.55 0.8   3.65    3.61 0.87 0.76   5  4.24 -0.6     0.01 0.04

stat.desc(props$Aud_eng)

##      nbr.val     nbr.null       nbr.na          min          max
##  343.0000000    7.0000000    0.0000000    0.0000000    5.0000000
##        range          sum       median         mean      SE.mean
##    5.0000000  592.3040000    1.0480000    1.7268338    0.0900376
## CI.mean.0.95          var      std.dev     coef.var
##    0.1770972    2.7806219    1.6675197    0.9656515

stat.desc(props$Hap_eng)

##      nbr.val     nbr.null       nbr.na          min          max
## 343.00000000   2.00000000   0.00000000   0.00000000   4.95200000
##        range          sum       median         mean      SE.mean
##   4.95200000 828.33800000   2.52400000   2.41497959   0.08728861
## CI.mean.0.95          var      std.dev     coef.var
##   0.17169012   2.61342058   1.61660774   0.66940845

stat.desc(props$Vis_eng)
```

```
##       nbr.val      nbr.null       nbr.na          min          max
## 3.430000e+02 0.000000e+00 0.000000e+00 5.240000e-01 5.000000e+00
##         range          sum       median         mean      SE.mean
## 4.476000e+00 1.301717e+03 4.167000e+00 3.795093e+00 5.737112e-02
## CI.mean.0.95          var      std.dev     coef.var
## 1.128447e-01 1.128966e+00 1.062528e+00 2.799741e-01

stat.desc(concs$Aud_eng)

##       nbr.val      nbr.null       nbr.na          min          max
## 392.00000000   1.00000000  24.00000000   0.00000000   5.00000000
##         range          sum       median         mean      SE.mean
##   5.00000000 848.38600000   2.05900000   2.16425000   0.05501011
## CI.mean.0.95          var      std.dev     coef.var
##   0.10815262   1.18623618   1.08914470   0.50324348

stat.desc(concs$Hap_eng)

##       nbr.val      nbr.null       nbr.na          min          max
##  392.0000000    1.0000000   24.0000000    0.0000000    4.7650000
##         range          sum       median         mean      SE.mean
##    4.7650000  730.5720000    1.6470000    1.8637041    0.0569453
## CI.mean.0.95          var      std.dev     coef.var
##    0.1119573    1.2711649    1.1274595    0.6049563

stat.desc(concs$Vis_eng)

##       nbr.val      nbr.null       nbr.na          min          max
## 3.920000e+02 0.000000e+00 2.400000e+01 7.650000e-01 5.000000e+00
##         range          sum       median         mean      SE.mean
## 4.235000e+00 1.393415e+03 3.647000e+00 3.554630e+00 4.053360e-02
## CI.mean.0.95          var      std.dev     coef.var
## 7.969106e-02 6.440452e-01 8.025243e-01 2.257687e-01

# Dutch
psych::describe(props$Auditory)

##    vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 336 1.74 1.29   1.33    1.55 0.96   0   5     5 1.14     0.26 0.07

psych::describe(props$Haptic)

##    vars   n mean   sd median trimmed  mad min  max range skew kurtosis
## X1    1 336 1.96 1.12   1.75    1.88 1.11   0 4.75  4.75 0.54    -0.58
##      se
## X1 0.06

psych::describe(props$Visual)

##    vars   n mean   sd median trimmed  mad min max range  skew kurtosis
## X1    1 336 3.22 1.15    3.4    3.31 1.26 0.5   5   4.5 -0.53    -0.68
```

```
##      se
## X1 0.06

psych::describe(concs$Auditory)

##    vars   n mean   sd median trimmed  mad min  max range skew kurtosis
## X1    1 411 1.97 1.03   1.86     1.9 1.04   0 4.89  4.89 0.59     0.16
##      se
## X1 0.05

psych::describe(concs$Haptic)

##    vars   n mean   sd median trimmed  mad min  max range skew kurtosis
## X1    1 411 1.96 1.04   1.78    1.87 0.99   0 4.78  4.78  0.7    -0.25
##      se
## X1 0.05

psych::describe(concs$Visual)

##    vars   n mean   sd median trimmed  mad  min max range skew kurtosis
## X1    1 411 3.13 0.95   3.22    3.16 0.99 0.33   5  4.67 -0.3    -0.49
##      se
## X1 0.05

stat.desc(props$Auditory)

##      nbr.val     nbr.null       nbr.na          min          max
## 336.00000000   3.00000000   7.00000000   0.00000000   5.00000000
##        range          sum       median         mean      SE.mean
##   5.00000000 584.54500000   1.33300000   1.73971726   0.07061977
## CI.mean.0.95          var      std.dev     coef.var
##   0.13891408   1.67568307   1.29448178   0.74407595

stat.desc(props$Haptic)

##      nbr.val     nbr.null       nbr.na          min          max
## 336.00000000   1.00000000   7.00000000   0.00000000   4.75000000
##        range          sum       median         mean      SE.mean
##   4.75000000 657.73300000   1.75000000   1.95753869   0.06113725
## CI.mean.0.95          var      std.dev     coef.var
##   0.12026128   1.25588834   1.12066424   0.57248638

stat.desc(props$Visual)

##      nbr.val     nbr.null       nbr.na          min          max
## 3.360000e+02 0.000000e+00 7.000000e+00 5.000000e-01 5.000000e+00
##        range          sum       median         mean      SE.mean
## 4.500000e+00 1.083009e+03 3.400000e+00 3.223241e+00 6.296073e-02
## CI.mean.0.95          var      std.dev     coef.var
## 1.238482e-01 1.331922e+00 1.154089e+00 3.580524e-01

stat.desc(concs$Auditory)
```

```
##      nbr.val     nbr.null       nbr.na          min          max
## 411.00000000   1.00000000   5.00000000   0.00000000   4.88900000
##        range          sum       median         mean      SE.mean
##   4.88900000 808.11100000   1.85700000   1.96620681   0.05082869
## CI.mean.0.95          var      std.dev     coef.var
##   0.09991736   1.06184159   1.03045698   0.52408372
```

```
stat.desc(concs$Haptic)
```

```
##      nbr.val     nbr.null       nbr.na          min          max
## 411.00000000   2.00000000   5.00000000   0.00000000   4.77800000
##        range          sum       median         mean      SE.mean
##   4.77800000 807.03000000   1.77800000   1.96357664   0.05141212
## CI.mean.0.95          var      std.dev     coef.var
##   0.10106424   1.08635771   1.04228485   0.53080935
```

```
stat.desc(concs$Visual)
```

```
##      nbr.val     nbr.null       nbr.na          min          max
## 4.110000e+02 0.000000e+00 5.000000e+00 3.330000e-01 5.000000e+00
##        range          sum       median         mean      SE.mean
## 4.667000e+00 1.285371e+03 3.222000e+00 3.127423e+00 4.695101e-02
## CI.mean.0.95          var      std.dev     coef.var
## 9.229475e-02 9.060075e-01 9.518442e-01 3.043541e-01
```

```
# Sample sizes for English and Dutch
nrow(props[!is.na(props$exc_eng),]) # total items w/ English norms = 343
```

```
## [1] 343
```

```
nrow(props[props$main_eng=='a' & !is.na(props$exc_eng),])
```

```
## [1] 68
```

```
nrow(props[props$main_eng=='h' & !is.na(props$exc_eng),])
```

```
## [1] 70
```

```
nrow(props[props$main_eng=='v' & !is.na(props$exc_eng),])
```

```
## [1] 205
```

```
nrow(props[!is.na(concs$exc_eng),]) # total items w/ English norms = 392
```

```
## [1] 392
```

```
nrow(props[concs$main_eng=='a' & !is.na(concs$exc_eng),])
```

```
## [1] 42
```

```
nrow(props[concs$main_eng=='h' & !is.na(concs$exc_eng),])
```

```
## [1] 14
```

```r
nrow(props[concs$main_eng=='v' & !is.na(concs$exc_eng),])

## [1] 336

nrow(props[!is.na(props$Exclusivity),]) # total props w/ Dutch norms = 336

## [1] 336

nrow(props[props$main=='a' & !is.na(props$Exclusivity),])

## [1] 64

nrow(props[props$main=='h' & !is.na(props$Exclusivity),])

## [1] 45

nrow(props[props$main=='v' & !is.na(props$Exclusivity),])

## [1] 227

nrow(props[!is.na(concs$Exclusivity),]) # total props w/ Dutch norms = 411

## [1] 411

nrow(props[concs$main=='a' & !is.na(concs$Exclusivity),])

## [1] 48

nrow(props[concs$main=='h' & !is.na(concs$Exclusivity),])

## [1] 45

nrow(props[concs$main=='v' & !is.na(concs$Exclusivity),])

## [1] 318

# ________________________________________________________________


# CRITICAL RESULTS: not translation-influenced

# Relation between modality strength, dominant modalities, and mod exclusivitY
# ENGLISH
# properties
summaryBy(Aud_eng ~ main_eng, data=props, FUN=mean)

##   main_eng Aud_eng.mean
## 1        a    4.5887941
## 2        h    1.1219429
## 3        v    0.9840488

summaryBy(Hap_eng ~ main_eng, data=props, FUN=mean)
```

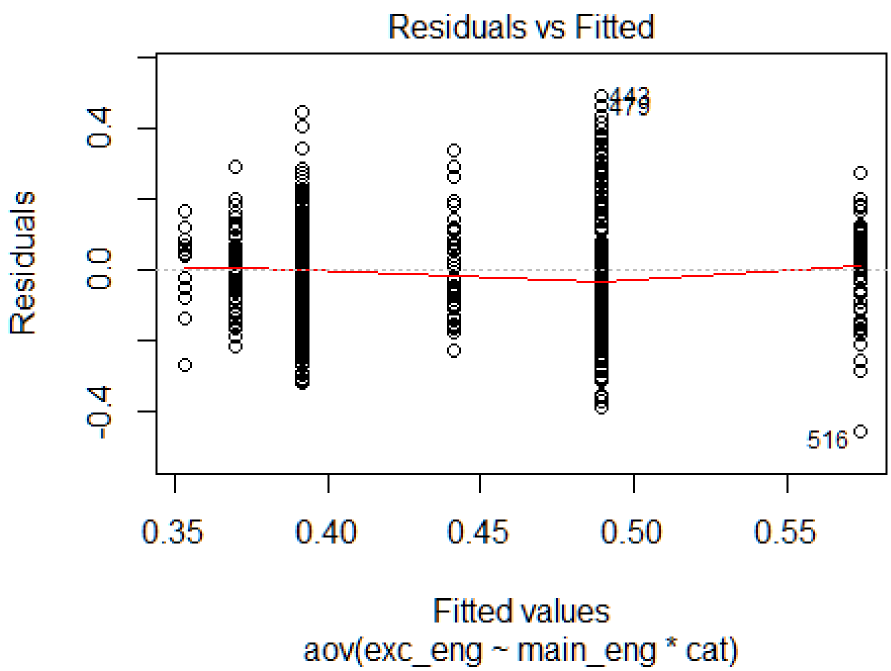
```
##   main_eng Hap_eng.mean
## 1        a    0.7042941
## 2        h    4.3319143
## 3        v    2.3278634

summaryBy(Vis_eng ~ main_eng, data=props, FUN=mean)

##   main_eng Vis_eng.mean
## 1        a     2.305074
## 2        h     3.447314
## 3        v     4.408098

summaryBy(exc_eng ~ main_eng, data=props, FUN=mean)

##   main_eng exc_eng.mean
## 1        a    0.5739265
## 2        h    0.3701571
## 3        v    0.4891659

# concepts
summaryBy(Aud_eng ~ main_eng, data=concs, FUN=mean)

##   main_eng Aud_eng.mean
## 1        a     3.542810
## 2        h     1.347643
## 3        v     2.025955
## 4     <NA>           NA

summaryBy(Hap_eng ~ main_eng, data=concs, FUN=mean)

##   main_eng Hap_eng.mean
## 1        a     1.032119
## 2        h     4.143143
## 3        v     1.872676
## 4     <NA>           NA

summaryBy(Vis_eng ~ main_eng, data=concs, FUN=mean)

##   main_eng Vis_eng.mean
## 1        a     2.711643
## 2        h     3.428714
## 3        v     3.665250
## 4     <NA>           NA

summaryBy(exc_eng ~ main_eng, data=concs, FUN=mean)

##   main_eng exc_eng.mean
## 1        a    0.4413571
## 2        h    0.3530714
## 3        v    0.3917173
## 4     <NA>           NA
```

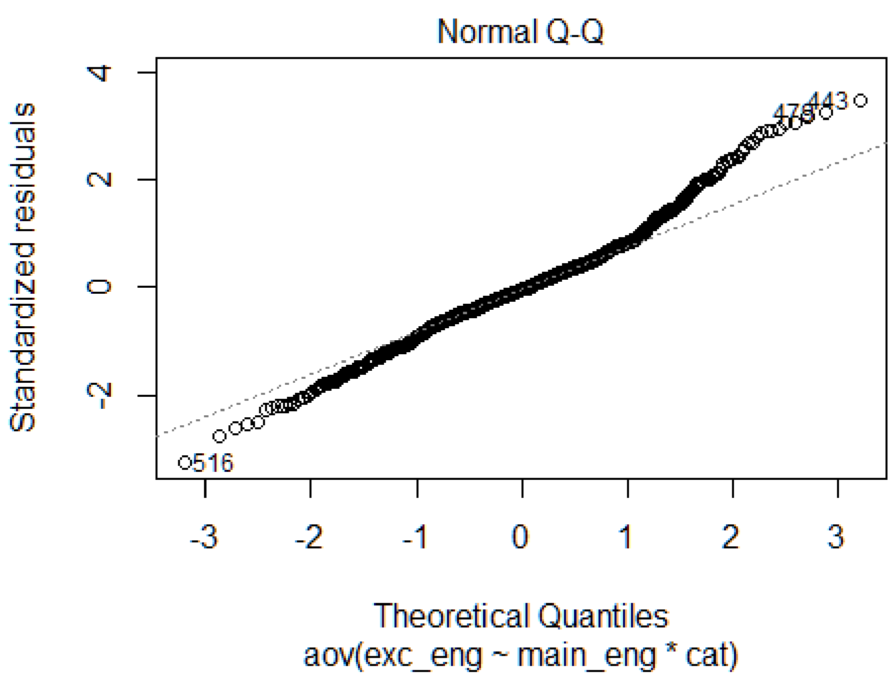
```r
# DUTCH
# properties
summaryBy(Auditory ~ main, data=props, FUN=mean)

##   main Auditory.mean
## 1    a      3.816750
## 2    h      1.374711
## 3    v      1.226480
## 4 <NA>            NA

summaryBy(Haptic ~ main, data=props, FUN=mean)

##   main Haptic.mean
## 1    a    1.220469
## 2    h    3.545978
## 3    v    1.850458
## 4 <NA>          NA

summaryBy(Visual ~ main, data=props, FUN=mean)

##   main Visual.mean
## 1    a    1.704125
## 2    h    2.722356
## 3    v    3.750833
## 4 <NA>          NA

summaryBy(Exclusivity ~ main, data=props, FUN=mean)

##   main Exclusivity.mean
## 1    a        0.4284531
## 2    h        0.2922667
## 3    v        0.4124714
## 4 <NA>               NA

# concepts
summaryBy(Auditory ~ main, data=concs, FUN=mean)

##   main Auditory.mean
## 1    a      3.447500
## 2    h      1.520956
## 3    v      1.805623
## 4 <NA>            NA

summaryBy(Haptic ~ main, data=concs, FUN=mean)

##   main Haptic.mean
## 1    a    1.498063
## 2    h    3.341511
## 3    v    1.838852
## 4 <NA>          NA

summaryBy(Visual ~ main, data=concs, FUN=mean)
```

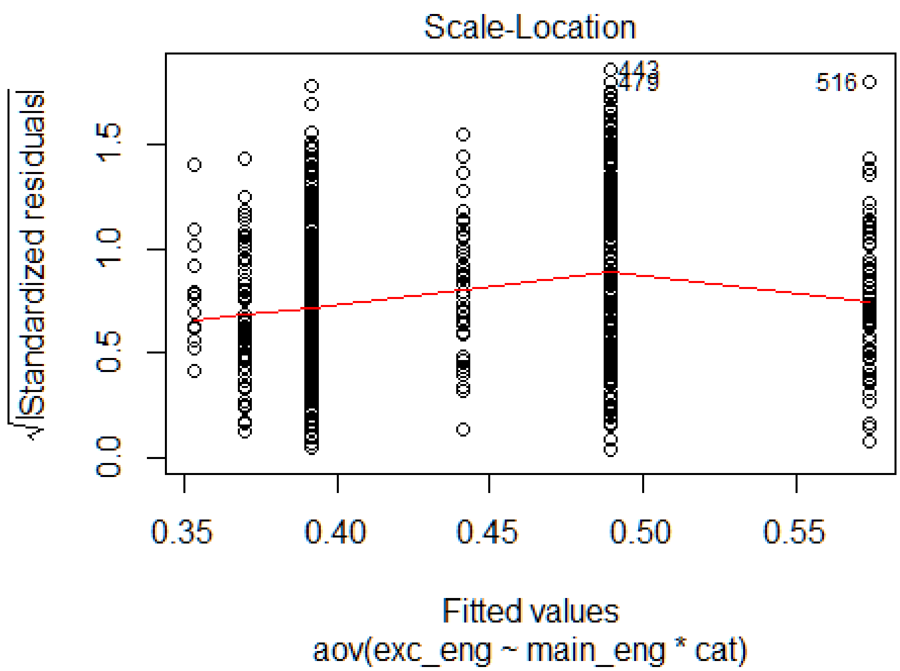
```
##   main Visual.mean
## 1    a    2.382125
## 2    h    2.721467
## 3    v    3.297368
## 4 <NA>          NA

summaryBy(Exclusivity ~ main, data=concs, FUN=mean)

##   main Exclusivity.mean
## 1    a        0.2806875
## 2    h        0.2571778
## 3    v        0.2905597
## 4 <NA>               NA

# RESULTS: both languages strongly related on individual modalities and on
# exclusivity. Correlations among modalities replicate previous norms, with visual
# and haptic items related, and auditory ones independent.

# Yet, there is clearly a greater exclusivity in the English norms.
# Properties
psych::describe(props$exc_eng)      # M = 0.48

##    vars   n mean   sd median trimmed  mad min  max range skew kurtosis
## X1    1 343 0.48 0.17   0.46    0.47 0.16 0.1 0.98  0.88 0.46    -0.11
##      se
## X1 0.01

psych::describe(props$Exclusivity)  # M = 0.40

##    vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 336  0.4 0.18   0.38    0.39 0.18   0   1     1 0.42    -0.22 0.01

# Concepts
psych::describe(concs$exc_eng)      # M = 0.40

##    vars   n mean   sd median trimmed mad  min  max range skew kurtosis
## X1    1 392  0.4 0.12   0.39    0.39 0.1 0.07 0.84  0.77  0.2      0.8
##      se
## X1 0.01

psych::describe(concs$Exclusivity)  # M = 0.29

##    vars   n mean   sd median trimmed  mad  min  max range skew kurtosis
## X1    1 411 0.29 0.15   0.27    0.28 0.15 0.01 0.91   0.9 0.62     0.38
##      se
## X1 0.01

# Indeed lower exclusivity and higher SD for Dutch items >> Check significance
# Because the English and the Dutch norms are paired, the difference has to be
# checked through a one-sample t-test, checking the mean of one language against
# the other language (see further below).
```

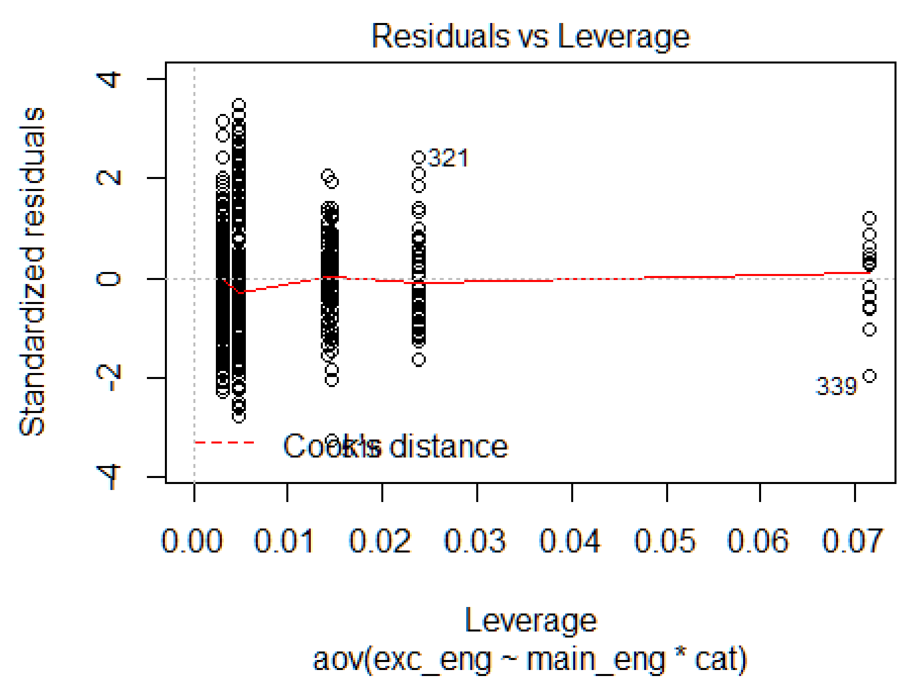
```r
# Correlations among modalities within each category and language:

# ENGLISH
# PROPERTIES
rcor.test(props[, c('Aud_eng', 'Hap_eng', 'Vis_eng', 'exc_eng')], use =
'complete.obs')

##
##         Aud_eng Hap_eng Vis_eng exc_eng
## Aud_eng  *****  -0.417  -0.625   0.018
## Hap_eng <0.001   *****   0.234  -0.621
## Vis_eng <0.001  <0.001   *****  -0.053
## exc_eng  0.740  <0.001   0.331   *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

corr3 = rcor.test(props[, c('Aud_eng', 'Hap_eng', 'Vis_eng', 'exc_eng')],
use = 'complete.obs')
write.csv(corr3$cor.mat, file = "corr3.csv",na="") # saved for manuscript

# CONCEPTS
rcor.test(concs[, c('Aud_eng', 'Hap_eng', 'Vis_eng', 'exc_eng')], use = 'complete.obs')

##
##         Aud_eng Hap_eng Vis_eng exc_eng
## Aud_eng  *****  -0.176  -0.008  -0.276
## Hap_eng <0.001   *****   0.554  -0.393
## Vis_eng  0.868  <0.001   *****  -0.065
## exc_eng <0.001  <0.001   0.202   *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

corr4 = rcor.test(concs[, c('Aud_eng', 'Hap_eng', 'Vis_eng', 'exc_eng')], use =
'complete.obs')
write.csv(corr4$cor.mat, file = "corr4.csv",na="") # saved for manuscript


# DUTCH
# PROPERTIES
rcor.test(props[, c('Auditory', 'Haptic', 'Visual', 'Exclusivity')], use =
'complete.obs')

##
##              Auditory Haptic Visual Exclusivity
## Auditory      *****   -0.228 -0.513 -0.173
## Haptic       <0.001    *****  0.193 -0.482
```

```
## Visual      <0.001   <0.001  *****  0.162
## Exclusivity  0.001   <0.001  0.003  *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

corr1 = rcor.test(props[, c('Auditory', 'Haptic', 'Visual', 'Exclusivity')],
use = 'complete.obs')
write.csv(corr1$cor.mat, file = "corr1.csv",na="") # saved for manuscript

# CONCEPTS
rcor.test(concs[, c('Auditory', 'Haptic', 'Visual', 'Exclusivity')], use =
'complete.obs')

##
##             Auditory Haptic Visual Exclusivity
## Auditory     *****   -0.009  0.085 -0.410
## Haptic       0.863    *****  0.441 -0.317
## Visual       0.086   <0.001  *****  0.122
## Exclusivity <0.001   <0.001  0.013  *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

corr2 = rcor.test(concs[, c('Auditory', 'Haptic', 'Visual', 'Exclusivity')], use =
'complete.obs')
write.csv(corr2$cor.mat, file = "corr2.csv",na="") # saved for manuscript


# Statistical tests for those differences
# Yet the same again, but now with a statistical significance test

# ENGLISH
# Setting contrasts based on means
contrasts(all$main_eng) <- cbind(c(2,0,-2), c(-1,2,-1))
# (1) Aud vs Vis; (2) Hap vs Aud-&-Vis
contrasts(all$main_eng)

##   [,1] [,2]
## a    2   -1
## h    0    2
## v   -2   -1

fitt <- aov(exc_eng ~ main_eng * cat, data=all)
plot(fitt)
```


Residuals vs Fitted
Residuals
0.4
0.0
-0.4
443
479
516
0.35
0.40
0.45
0.50
0.55
Fitted values
aov(exc_eng ~ main_eng * cat)


Normal Q-Q
Standardized residuals
4
2
0
-2
479
443
516
-3
-2
-1
0
1
2
3
Theoretical Quantiles
aov(exc_eng ~ main_eng * cat)






```
summary(fitt)

##                 Df Sum Sq Mean Sq F value   Pr(>F)
## main_eng         2  1.264  0.6321  31.507 7.52e-14 ***
```

```
## cat            1  1.561  1.5614  77.832  < 2e-16 ***
## main_eng:cat   2  0.107  0.0537   2.676   0.0695 .
## Residuals    729 14.625  0.0201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 24 observations deleted due to missingness
```

```
drop1(fitt,~.,test="F")
```

```
## Single term deletions
##
## Model:
## exc_eng ~ main_eng * cat
##              Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>                    14.624 -2867.1
## main_eng      2   0.11832 14.743 -2865.2   2.949   0.05302 .
## cat           1   0.46228 15.087 -2846.2  23.044 1.922e-06 ***
## main_eng:cat  2   0.10737 14.732 -2865.8   2.676   0.06951 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(fitt)
```

```
## Anova Table (Type II tests)
##
## Response: exc_eng
##               Sum Sq  Df F value    Pr(>F)
## main_eng      1.4717   2  36.681 6.628e-16 ***
## cat           1.5614   1  77.832 < 2.2e-16 ***
## main_eng:cat  0.1074   2   2.676   0.06951 .
## Residuals    14.6245 729
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(fitt, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: exc_eng
##               Sum Sq  Df F value    Pr(>F)
## main_eng      1.4717   2  36.681 6.628e-16 ***
## cat           1.5614   1  77.832 < 2.2e-16 ***
## main_eng:cat  0.1074   2   2.676   0.06951 .
## Residuals    14.6245 729
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(fitt, type = "III")
```

```
## Anova Table (Type III tests)
##
## Response: exc_eng
```

```
##                Sum Sq  Df F value    Pr(>F)
## (Intercept)  14.3252   1 714.082 < 2.2e-16 ***
## main_eng      0.1183   2   2.949   0.05302 .
## cat           0.4623   1  23.044 1.922e-06 ***
## main_eng:cat  0.1074   2   2.676   0.06951 .
## Residuals    14.6245 729
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

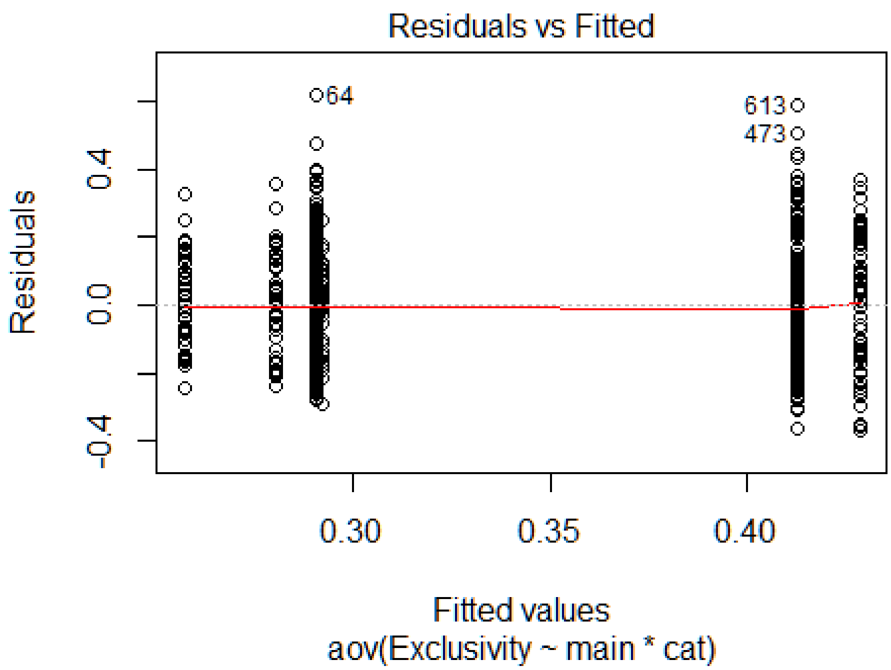
```r
summary.lm(fitt)
```

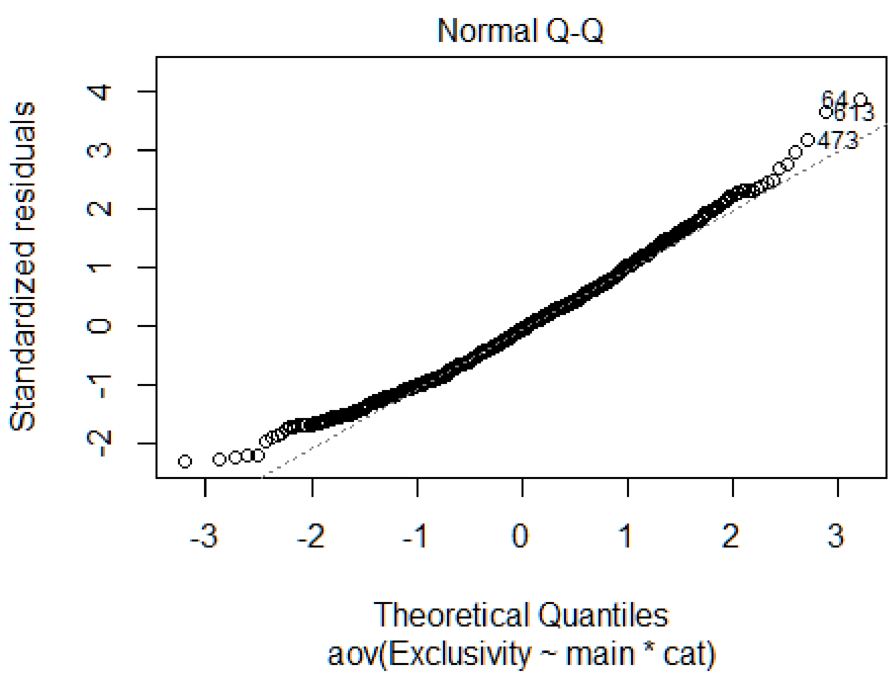
```
##
## Call:
## aov(formula = exc_eng ~ main_eng * cat, data = all)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45793 -0.08089 -0.00472  0.06968  0.49083
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.395382   0.014796  26.722  < 2e-16 ***
## main_eng1           0.012410   0.005795   2.141   0.0326 *
## main_eng2          -0.021155   0.013196  -1.603   0.1093
## catprop             0.082368   0.017159   4.800 1.92e-06 ***
## main_eng1:catprop   0.008780   0.007625   1.152   0.2499
## main_eng2:catprop  -0.032641   0.014727  -2.216   0.0270 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1416 on 729 degrees of freedom
##   (24 observations deleted due to missingness)
## Multiple R-squared:  0.167,  Adjusted R-squared:  0.1613
## F-statistic: 29.24 on 5 and 729 DF,  p-value: < 2.2e-16
```

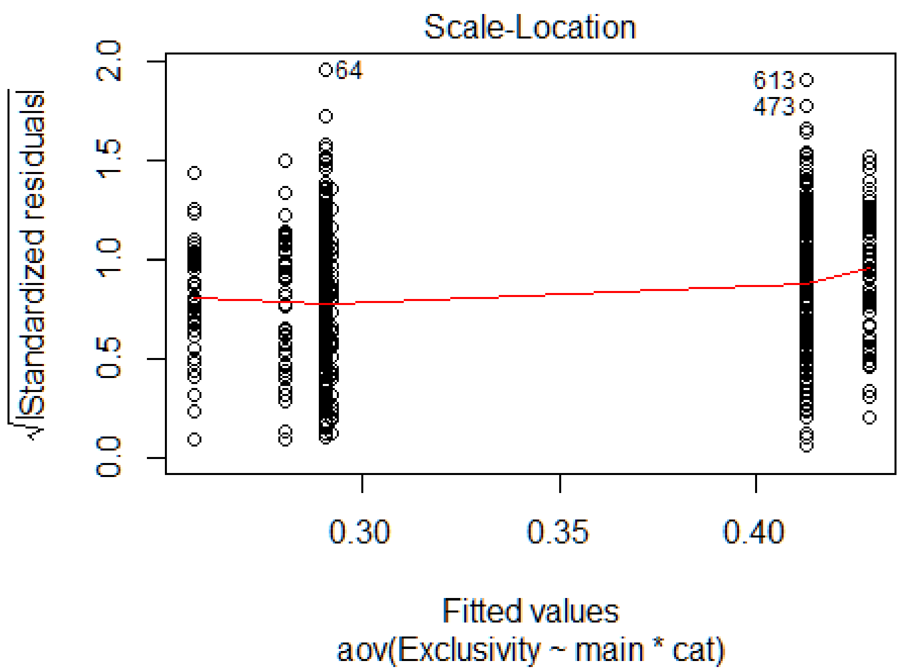
```r
# RESULTS: English properties with more exclusivity than concepts(***)
# Contrasts: (1) Aud vs Vis (*)
# (2) Haptic words show less exclusivity than auditory and visual ones within
# properties but not within concepts (*)


# DUTCH
# Setting contrasts based on means
contrasts(all$main) <- cbind(c(2,0,-2), c(-1,2,-1))
# (1) Aud vs Vis; (2) Hap vs Aud-&-Vis
contrasts(all$main)
```

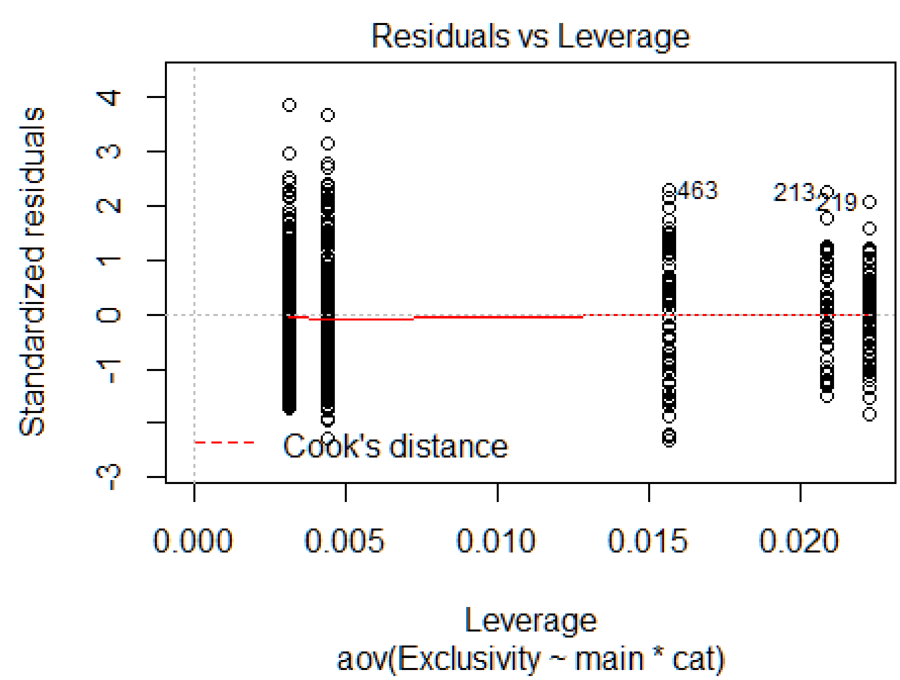
```
##   [,1] [,2]
## a    2   -1
## h    0    2
## v   -2   -1
```

```
fitt <- aov(Exclusivity ~ main * cat, data=all)
plot(fitt)  # must click over the plot several times in order to continue
```










```
summary(fitt)

##                Df Sum Sq Mean Sq F value   Pr(>F)
## main            2  0.448  0.2240   8.679 0.000188 ***
```

```
## cat           1  2.416  2.4157  93.616  < 2e-16 ***
## main:cat      2  0.179  0.0897   3.477 0.031399 *
## Residuals   741 19.121  0.0258
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 12 observations deleted due to missingness
```

```
drop1(fitt,~.,test="F")
```

```
## Single term deletions
##
## Model:
## Exclusivity ~ main * cat
##          Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>                19.121 -2726.0
## main      2   0.04532 19.166 -2728.2  0.8782    0.4160
## cat       1   1.05008 20.171 -2688.0 40.6939 3.134e-10 ***
## main:cat  2   0.17945 19.300 -2723.0  3.4772    0.0314 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(fitt)
```

```
## Anova Table (Type II tests)
##
## Response: Exclusivity
##            Sum Sq  Df F value    Pr(>F)
## main       0.4752   2  9.2071 0.0001123 ***
## cat        2.4157   1 93.6158 < 2.2e-16 ***
## main:cat   0.1795   2  3.4772 0.0313993 *
## Residuals 19.1210 741
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(fitt, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: Exclusivity
##            Sum Sq  Df F value    Pr(>F)
## main       0.4752   2  9.2071 0.0001123 ***
## cat        2.4157   1 93.6158 < 2.2e-16 ***
## main:cat   0.1795   2  3.4772 0.0313993 *
## Residuals 19.1210 741
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(fitt, type = "III")
```

```
## Anova Table (Type III tests)
##
## Response: Exclusivity
```

```
##              Sum Sq  Df  F value    Pr(>F)
## (Intercept) 14.8547   1 575.6647 < 2.2e-16 ***
## main         0.0453   2   0.8782    0.4160
## cat          1.0501   1  40.6939 3.134e-10 ***
## main:cat     0.1795   2   3.4772    0.0314 *
## Residuals   19.1210 741
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary.lm(fitt)

##
## Call:
## aov(formula = Exclusivity ~ main * cat, data = all)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36745 -0.11897 -0.00947  0.09959  0.61844
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.276142   0.011509  23.993  < 2e-16 ***
## main1          -0.002468   0.006219  -0.397  0.69157
## main2          -0.009482   0.008995  -1.054  0.29214
## catprop         0.101589   0.015925   6.379 3.13e-10 ***
## main1:catprop  0.006464   0.008425   0.767  0.44320
## main2:catprop -0.033250   0.012608  -2.637  0.00854 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1606 on 741 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.1373, Adjusted R-squared:  0.1315
## F-statistic: 23.59 on 5 and 741 DF,  p-value: < 2.2e-16

# RESULTS: Dutch properties with more exclusivity than concepts(***)
# Contrasts: (1) Aud vs Vis (non-sig)
# (2) Haptic words show less exclusivity than auditory and visual ones within
# properties but not within concepts (**)

# Overall, these results stem from the nature of human perception. What exclusivity
# seems to be indexing is the degree to which percepts will naturally co-occur. Thus,
# visual and auditory words have relatively higher exclusivities because what we
# see and hear often stands on its own. We can often see thing but not hear or touch
# them. 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 and hear it too--hence the
# low exclusivity of haptic items.
```

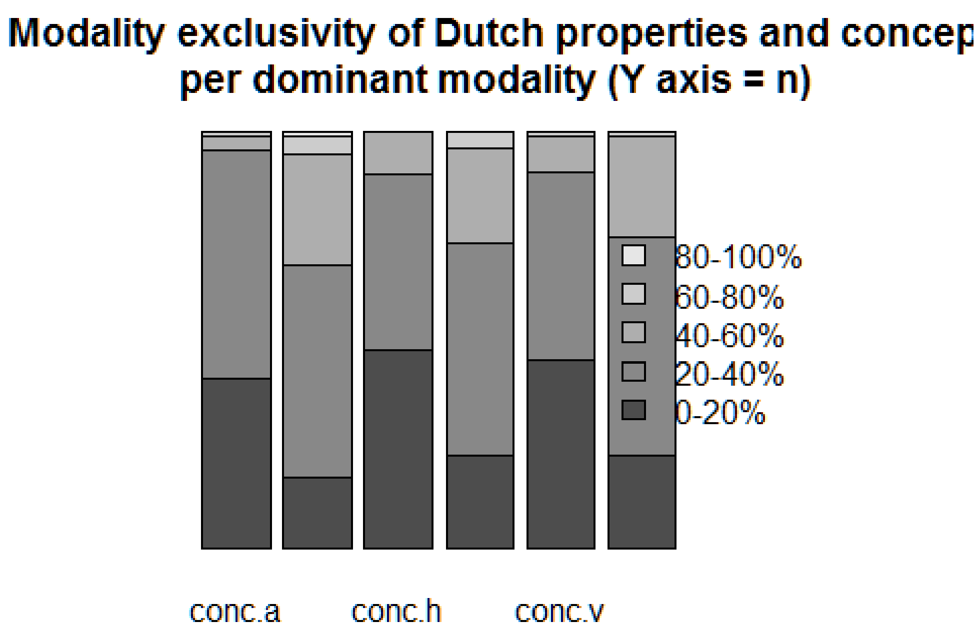
```
# SAME PLOT-WISE:
# Barplot of exclusivity percentiles within modalities for Dutch items (as in
# van Dantzig et al., 2011, but separately for properties and concepts)

all<-read.csv('all.csv')

allNL = all[!all$main == '',]
allNL$main = levels(droplevels(allNL$main))

concs <- allNL[allNL$cat == 'conc' & !allNL$normed == 'English' & !allNL$main == '',]
props <- allNL[allNL$cat == 'prop' & !allNL$normed == 'English' & !allNL$main == '',]

concs$main = levels(droplevels(as.factor(concs$main)))
props$main = levels(droplevels(as.factor(props$main)))
concs$main = as.factor(concs$main)
props$main = as.factor(props$main)
nrow(concs$main)

## NULL

nrow(props$main)

## NULL

allNL$catmain <- with(allNL, interaction(cat,  main))
str(allNL$catmain)

##  Factor w/ 6 levels "conc.a","prop.a",..: 1 3 5 1 3 5 1 3 5 1 ...

allNL$section = floor(allNL$Exclusivity * 4)
table(allNL$section)

##
##   0   1   2   3   4
## 252 368 114  12   1

str(allNL$section)

##  num [1:759] 1 0 1 1 2 1 2 2 1 1 ...

table(allNL$section)  # order = 01234

##
##   0   1   2   3   4
## 252 368 114  12   1

allNL$section = as.factor(allNL$section)
revalue(allNL$section, c("0"="0-20%", "1"="20-40%", "2"="40-60%", "3"="60-80%",
"4"="80-100%"))

##   [1] 20-40%  0-20%   20-40%  20-40%  40-60%  20-40%  40-60%  40-60%
##   [9] 20-40%  20-40%  0-20%   0-20%   0-20%   0-20%   20-40%  0-20%
```

```
##  [17] 0-20%   20-40%  20-40%  0-20%   20-40%  0-20%   20-40%  20-40%
##  [25] 0-20%   40-60%  0-20%   0-20%   0-20%   0-20%   20-40%  40-60%
##  [33] 0-20%   20-40%  20-40%  20-40%  20-40%  0-20%   0-20%   20-40%
##  [41] 0-20%   0-20%   0-20%   0-20%   0-20%   0-20%   20-40%  0-20%
##  [49] 20-40%  40-60%  0-20%   0-20%   40-60%  0-20%   20-40%  40-60%
##  [57] 0-20%   20-40%  0-20%   0-20%   20-40%  20-40%  20-40%  60-80%
##  [65] 0-20%   0-20%   20-40%  0-20%   40-60%  0-20%   20-40%  0-20%
##  [73] 20-40%  0-20%   20-40%  0-20%   20-40%  0-20%   20-40%  0-20%
##  [81] 0-20%   0-20%   40-60%  0-20%   0-20%   0-20%   40-60%  20-40%
##  [89] 0-20%   0-20%   0-20%   20-40%  20-40%  0-20%   20-40%  20-40%
##  [97] 0-20%   20-40%  0-20%   20-40%  20-40%  0-20%   0-20%   20-40%
## [105] 0-20%   20-40%  20-40%  0-20%   0-20%   0-20%   0-20%   0-20%
## [113] 20-40%  0-20%   0-20%   0-20%   20-40%  0-20%   0-20%   0-20%
## [121] 20-40%  0-20%   0-20%   0-20%   20-40%  40-60%  20-40%  0-20%
## [129] 20-40%  0-20%   0-20%   0-20%   20-40%  20-40%  20-40%  20-40%
## [137] 20-40%  0-20%   20-40%  0-20%   20-40%  20-40%  20-40%  0-20%
## [145] 0-20%   0-20%   0-20%   0-20%   40-60%  40-60%  0-20%   0-20%
## [153] 20-40%  20-40%  0-20%   20-40%  0-20%   20-40%  0-20%   0-20%
## [161] 0-20%   20-40%  20-40%  20-40%  0-20%   20-40%  0-20%   0-20%
## [169] 20-40%  20-40%  0-20%   0-20%   20-40%  20-40%  40-60%  20-40%
## [177] 20-40%  20-40%  20-40%  40-60%  20-40%  0-20%   20-40%  20-40%
## [185] 20-40%  40-60%  20-40%  0-20%   20-40%  20-40%  0-20%   0-20%
## [193] 0-20%   0-20%   0-20%   0-20%   0-20%   0-20%   0-20%   0-20%
## [201] 20-40%  20-40%  20-40%  20-40%  0-20%   20-40%  0-20%   20-40%
## [209] 0-20%   20-40%  20-40%  20-40%  40-60%  0-20%   20-40%  20-40%
## [217] 20-40%  0-20%   40-60%  0-20%   20-40%  20-40%  20-40%  20-40%
## [225] 0-20%   20-40%  20-40%  20-40%  20-40%  0-20%   20-40%  20-40%
## [233] 20-40%  20-40%  0-20%   20-40%  20-40%  20-40%  0-20%   0-20%
## [241] 20-40%  20-40%  20-40%  20-40%  20-40%  20-40%  20-40%  0-20%
## [249] 20-40%  20-40%  0-20%   20-40%  0-20%   20-40%  20-40%  0-20%
## [257] 0-20%   0-20%   0-20%   20-40%  20-40%  20-40%  0-20%   0-20%
## [265] 0-20%   0-20%   20-40%  0-20%   0-20%   0-20%   0-20%   0-20%
## [273] 0-20%   20-40%  20-40%  0-20%   20-40%  0-20%   20-40%  20-40%
## [281] 20-40%  20-40%  0-20%   20-40%  20-40%  20-40%  0-20%   0-20%
## [289] 20-40%  0-20%   20-40%  20-40%  20-40%  20-40%  0-20%   0-20%
## [297] 0-20%   20-40%  40-60%  20-40%  20-40%  20-40%  0-20%   20-40%
## [305] 20-40%  40-60%  20-40%  0-20%   20-40%  20-40%  0-20%   20-40%
## [313] 20-40%  0-20%   20-40%  20-40%  40-60%  20-40%  20-40%  0-20%
## [321] 40-60%  0-20%   0-20%   0-20%   20-40%  0-20%   0-20%   0-20%
## [329] 20-40%  0-20%   20-40%  20-40%  20-40%  0-20%   0-20%   20-40%
## [337] 0-20%   40-60%  20-40%  20-40%  40-60%  0-20%   20-40%  20-40%
## [345] 20-40%  0-20%   20-40%  0-20%   0-20%   0-20%   0-20%   0-20%
## [353] 20-40%  0-20%   20-40%  20-40%  0-20%   20-40%  0-20%   40-60%
## [361] 40-60%  20-40%  20-40%  20-40%  20-40%  20-40%  0-20%   20-40%
## [369] 20-40%  20-40%  20-40%  60-80%  0-20%   20-40%  20-40%  20-40%
## [377] 0-20%   20-40%  20-40%  40-60%  0-20%   0-20%   0-20%   0-20%
## [385] 0-20%   20-40%  40-60%  20-40%  0-20%   20-40%  40-60%  0-20%
## [393] 0-20%   20-40%  0-20%   20-40%  0-20%   20-40%  0-20%   20-40%
## [401] 20-40%  0-20%   20-40%  0-20%   0-20%   0-20%   20-40%  20-40%
## [409] 40-60%  0-20%   20-40%  0-20%   20-40%  20-40%  0-20%   20-40%
```

```
## [417] 40-60%  40-60%  20-40%  20-40%  40-60%  20-40%  40-60%  40-60%
## [425] 0-20%   40-60%  20-40%  20-40%  40-60%  20-40%  20-40%  20-40%
## [433] 0-20%   0-20%   20-40%  0-20%   0-20%   20-40%  20-40%  60-80%
## [441] 20-40%  40-60%  20-40%  0-20%   20-40%  20-40%  20-40%  20-40%
## [449] 20-40%  20-40%  20-40%  40-60%  20-40%  40-60%  20-40%  40-60%
## [457] 20-40%  0-20%   0-20%   40-60%  20-40%  0-20%   60-80%  20-40%
## [465] 20-40%  40-60%  20-40%  40-60%  40-60%  20-40%  20-40%  20-40%
## [473] 60-80%  40-60%  20-40%  20-40%  0-20%   40-60%  40-60%  40-60%
## [481] 0-20%   20-40%  0-20%   20-40%  20-40%  20-40%  60-80%  40-60%
## [489] 20-40%  40-60%  20-40%  0-20%   20-40%  20-40%  20-40%  20-40%
## [497] 20-40%  20-40%  40-60%  40-60%  20-40%  20-40%  60-80%  20-40%
## [505] 40-60%  40-60%  40-60%  40-60%  0-20%   20-40%  40-60%  20-40%
## [513] 40-60%  20-40%  40-60%  20-40%  20-40%  20-40%  60-80%  0-20%
## [521] 40-60%  20-40%  40-60%  0-20%   20-40%  20-40%  0-20%   40-60%
## [529] 60-80%  20-40%  0-20%   20-40%  20-40%  0-20%   20-40%  0-20%
## [537] 20-40%  20-40%  0-20%   0-20%   60-80%  20-40%  20-40%  40-60%
## [545] 20-40%  0-20%   20-40%  40-60%  20-40%  20-40%  20-40%  40-60%
## [553] 20-40%  0-20%   20-40%  20-40%  40-60%  0-20%   20-40%  20-40%
## [561] 20-40%  40-60%  20-40%  20-40%  0-20%   40-60%  0-20%   0-20%
## [569] 40-60%  40-60%  40-60%  20-40%  40-60%  20-40%  0-20%   20-40%
## [577] 40-60%  0-20%   20-40%  40-60%  40-60%  40-60%  20-40%  40-60%
## [585] 40-60%  20-40%  20-40%  40-60%  0-20%   40-60%  20-40%  20-40%
## [593] 20-40%  20-40%  40-60%  40-60%  0-20%   40-60%  20-40%  0-20%
## [601] <NA>    <NA>    <NA>    <NA>    <NA>    <NA>    <NA>    <NA>
## [609] <NA>    <NA>    <NA>    <NA>    80-100% 20-40%  0-20%   20-40%
## [617] 20-40%  20-40%  20-40%  20-40%  20-40%  20-40%  20-40%  40-60%
## [625] 20-40%  40-60%  40-60%  60-80%  0-20%   20-40%  20-40%  40-60%
## [633] 0-20%   0-20%   0-20%   20-40%  0-20%   20-40%  20-40%  20-40%
## [641] 0-20%   20-40%  0-20%   0-20%   20-40%  20-40%  40-60%  40-60%
## [649] 40-60%  0-20%   20-40%  20-40%  20-40%  0-20%   20-40%  40-60%
## [657] 40-60%  20-40%  20-40%  20-40%  40-60%  0-20%   0-20%   40-60%
## [665] 40-60%  20-40%  20-40%  0-20%   20-40%  20-40%  20-40%  0-20%
## [673] 40-60%  0-20%   20-40%  20-40%  20-40%  40-60%  0-20%   40-60%
## [681] 20-40%  20-40%  0-20%   20-40%  20-40%  20-40%  20-40%  0-20%
## [689] 20-40%  20-40%  20-40%  20-40%  20-40%  20-40%  20-40%  20-40%
## [697] 20-40%  0-20%   40-60%  20-40%  20-40%  0-20%   20-40%  20-40%
## [705] 20-40%  20-40%  20-40%  0-20%   20-40%  40-60%  0-20%   20-40%
## [713] 20-40%  20-40%  40-60%  0-20%   40-60%  40-60%  40-60%  20-40%
## [721] 0-20%   20-40%  20-40%  0-20%   20-40%  20-40%  20-40%  40-60%
## [729] 20-40%  40-60%  20-40%  0-20%   0-20%   0-20%   20-40%  20-40%
## [737] 20-40%  40-60%  40-60%  20-40%  0-20%   0-20%   60-80%  20-40%
## [745] 0-20%   0-20%   40-60%  20-40%  20-40%  20-40%  20-40%  40-60%
## [753] 40-60%  20-40%  20-40%  0-20%   20-40%  40-60%  20-40%
## Levels: 0-20% 20-40% 40-60% 60-80% 80-100%
```

```
allNL$section = mapvalues(allNL$section, from = c(0, 1, 2, 3, 4), to = c("0-20%",
"20-40%", "40-60%", "60-80%", "80-100%"))
table(allNL$section)
```

```
##
##    0-20%  20-40%  40-60%  60-80% 80-100%
##      252     368     114      12       1

str(allNL$section)

##  Factor w/ 5 levels "0-20%","20-40%",..: 2 1 2 2 3 2 3 3 2 2 ...

counts <- table(allNL$section, allNL$catmain)
counts

##
##             conc.a prop.a conc.h prop.h conc.v prop.v
##    0-20%       56     19     65     25     62     25
##    20-40%      75     57     58     57     62     59
##    40-60%       5     30     14     26     12     27
##    60-80%       1      5      0      4      1      1
##    80-100%      0      1      0      0      0      0

counts = prop.table(counts, 2)

# see plot:
barplot(counts, width=10, main = 'Modality exclusivity of Dutch properties and concepts
per dominant modality (Y axis = n)       ', legend = rownames(counts), xlim=c(0,100),
axes=FALSE, args.legend = list(x = "topright", bty = "n", inset=c(.1, .2)))
```




```
# ! THE PLOT IS SHOWN BADLY ON HERE. PLEASE SEE THE SAVED PLOT.
```

```r
# Below, run first line, then return and keep running:
png(file="stacked_exc.png", units="in", width=6, height=6, res=1000)
par(mar=c(2,-.3,3,-.3)+.4)  # run twice, if necessary
barplot(counts, width=10, main = 'Modality exclusivity of Dutch properties and concepts
per dominant modality (Y axis = n)      ', legend = rownames(counts), xlim=c(0,100),
axes=FALSE, args.legend = list(x = "topright", bty = "n", inset=c(.1, .2)))
dev.off()

## png
##   2

# Same plot for the English items of Lynott and Connell (of course w/out gustatory
# or olfactory)

allENG = all[!all$main_eng == '',]
allENG$main_eng = levels(droplevels(allENG$main_eng))

allENG$catmain <- with(allENG, interaction(cat,  main_eng))
str(allENG$catmain)

##  Factor w/ 6 levels "conc.a","prop.a",..: NA 3 5 1 3 5 1 3 5 1 ...

allENG$section = floor(allENG$exc_eng * 5)
table(allENG$section)

##
##   0   1   2   3   4
##  33 288 313  81  20

str(allENG$section)

##  num [1:759] NA 2 1 2 2 2 3 2 1 2 ...

table(allENG$section)  # order = 01234

##
##   0   1   2   3   4
##  33 288 313  81  20

allENG$section = as.factor(allENG$section)
revalue(allENG$section, c("0"="0-20%", "1"="20-40%", "2"="40-60%", "3"="60-80%",
"4"="80-100%"))

##   [1] <NA>    40-60%  20-40%  40-60%  40-60%  40-60%  60-80%  40-60%
##   [9] 20-40%  40-60%  40-60%  20-40%  40-60%  20-40%  40-60%  20-40%
##  [17] 20-40%  40-60%  40-60%  40-60%  20-40%  0-20%   20-40%  20-40%
##  [25] 60-80%  40-60%  <NA>    40-60%  40-60%  20-40%  60-80%  40-60%
##  [33] 20-40%  40-60%  40-60%  20-40%  20-40%  20-40%  20-40%  20-40%
##  [41] 20-40%  20-40%  20-40%  40-60%  20-40%  40-60%  20-40%  40-60%
##  [49] 40-60%  20-40%  40-60%  20-40%  20-40%  20-40%  20-40%  20-40%
##  [57] 20-40%  20-40%  20-40%  40-60%  40-60%  40-60%  20-40%  40-60%
##  [65] 40-60%  40-60%  40-60%  20-40%  40-60%  0-20%   20-40%  0-20%
```

```
##  [73] 20-40%  40-60%  <NA>    60-80%  40-60%  40-60%  40-60%  20-40% 
##  [81] 20-40%  20-40%  20-40%  20-40%  20-40%  20-40%  60-80%  40-60% 
##  [89] 20-40%  20-40%  20-40%  20-40%  0-20%   20-40%  40-60%  20-40% 
##  [97] 20-40%  40-60%  20-40%  20-40%  20-40%  40-60%  20-40%  20-40% 
## [105] 20-40%  20-40%  40-60%  20-40%  20-40%  20-40%  20-40%  <NA>   
## [113] <NA>    20-40%  20-40%  40-60%  40-60%  <NA>    0-20%   0-20%  
## [121] <NA>    20-40%  40-60%  20-40%  40-60%  40-60%  20-40%  20-40% 
## [129] 40-60%  <NA>    40-60%  40-60%  <NA>    40-60%  40-60%  40-60% 
## [137] 40-60%  20-40%  60-80%  0-20%   40-60%  60-80%  60-80%  40-60% 
## [145] 20-40%  40-60%  20-40%  20-40%  40-60%  40-60%  40-60%  20-40% 
## [153] <NA>    20-40%  20-40%  20-40%  40-60%  20-40%  20-40%  20-40% 
## [161] 20-40%  <NA>    40-60%  40-60%  40-60%  40-60%  40-60%  40-60% 
## [169] 40-60%  20-40%  40-60%  20-40%  40-60%  20-40%  40-60%  40-60% 
## [177] 40-60%  40-60%  40-60%  80-100% 40-60%  20-40%  60-80%  40-60% 
## [185] 40-60%  40-60%  40-60%  40-60%  40-60%  40-60%  40-60%  20-40% 
## [193] 20-40%  20-40%  20-40%  20-40%  20-40%  20-40%  40-60%  0-20%  
## [201] 40-60%  40-60%  <NA>    20-40%  40-60%  20-40%  40-60%  20-40% 
## [209] 20-40%  20-40%  <NA>    20-40%  40-60%  20-40%  <NA>    40-60% 
## [217] 40-60%  20-40%  <NA>    20-40%  20-40%  20-40%  40-60%  60-80% 
## [225] 20-40%  40-60%  0-20%   20-40%  40-60%  40-60%  20-40%  40-60% 
## [233] 40-60%  20-40%  <NA>    40-60%  40-60%  20-40%  20-40%  20-40% 
## [241] 40-60%  20-40%  20-40%  40-60%  20-40%  40-60%  40-60%  40-60% 
## [249] 20-40%  40-60%  0-20%   40-60%  <NA>    0-20%   20-40%  20-40% 
## [257] 20-40%  20-40%  20-40%  40-60%  60-80%  20-40%  20-40%  40-60% 
## [265] 40-60%  20-40%  40-60%  20-40%  <NA>    40-60%  40-60%  40-60% 
## [273] 0-20%   20-40%  <NA>    20-40%  <NA>    20-40%  60-80%  20-40% 
## [281] 20-40%  40-60%  20-40%  40-60%  20-40%  0-20%   20-40%  20-40% 
## [289] 40-60%  20-40%  40-60%  20-40%  40-60%  20-40%  40-60%  20-40% 
## [297] 20-40%  40-60%  40-60%  60-80%  <NA>    20-40%  20-40%  40-60% 
## [305] 60-80%  20-40%  40-60%  20-40%  40-60%  0-20%   20-40%  20-40% 
## [313] 20-40%  20-40%  40-60%  40-60%  40-60%  40-60%  40-60%  20-40% 
## [321] 60-80%  40-60%  0-20%   20-40%  <NA>    40-60%  40-60%  20-40% 
## [329] 40-60%  20-40%  20-40%  <NA>    40-60%  20-40%  0-20%   20-40% 
## [337] 40-60%  40-60%  0-20%   20-40%  40-60%  20-40%  20-40%  40-60% 
## [345] <NA>    0-20%   0-20%   40-60%  20-40%  20-40%  40-60%  20-40% 
## [353] 40-60%  20-40%  0-20%   40-60%  0-20%   20-40%  40-60%  60-80% 
## [361] 60-80%  20-40%  40-60%  40-60%  20-40%  40-60%  20-40%  40-60% 
## [369] 40-60%  20-40%  20-40%  20-40%  20-40%  40-60%  40-60%  20-40% 
## [377] 40-60%  20-40%  20-40%  40-60%  20-40%  20-40%  40-60%  20-40% 
## [385] 40-60%  0-20%   60-80%  40-60%  20-40%  20-40%  40-60%  20-40% 
## [393] 20-40%  40-60%  20-40%  40-60%  20-40%  40-60%  40-60%  40-60% 
## [401] 40-60%  20-40%  20-40%  20-40%  20-40%  20-40%  60-80%  40-60% 
## [409] 20-40%  20-40%  60-80%  20-40%  40-60%  60-80%  20-40%  40-60% 
## [417] 80-100% 20-40%  40-60%  20-40%  80-100% 40-60%  80-100% 80-100%
## [425] 20-40%  60-80%  40-60%  40-60%  80-100% 40-60%  40-60%  20-40% 
## [433] 40-60%  0-20%   40-60%  20-40%  40-60%  0-20%   40-60%  60-80% 
## [441] 40-60%  80-100% 80-100% 40-60%  40-60%  40-60%  40-60%  20-40% 
## [449] 20-40%  20-40%  40-60%  60-80%  40-60%  20-40%  40-60%  60-80% 
## [457] 40-60%  20-40%  20-40%  40-60%  20-40%  40-60%  80-100% 20-40% 
## [465] 20-40%  40-60%  40-60%  60-80%  40-60%  60-80%  60-80%  60-80%
```

```
## [473] 20-40%  60-80%  80-100% 40-60%  20-40%  40-60%  80-100% 60-80%
## [481] 20-40%  60-80%  20-40%  40-60%  40-60%  20-40%  60-80%  60-80%
## [489] 40-60%  40-60%  40-60%  40-60%  40-60%  40-60%  40-60%  40-60%
## [497] 20-40%  40-60%  40-60%  40-60%  60-80%  20-40%  40-60%  20-40%
## [505] 60-80%  60-80%  60-80%  60-80%  20-40%  40-60%  60-80%  20-40%
## [513] 80-100% 40-60%  60-80%  0-20%   40-60%  20-40%  0-20%   40-60%
## [521] 40-60%  20-40%  80-100% 20-40%  40-60%  40-60%  20-40%  40-60%
## [529] 60-80%  40-60%  40-60%  40-60%  20-40%  20-40%  40-60%  60-80%
## [537] 40-60%  40-60%  20-40%  20-40%  80-100% 40-60%  20-40%  40-60%
## [545] 40-60%  40-60%  40-60%  60-80%  40-60%  20-40%  20-40%  40-60%
## [553] 20-40%  0-20%   40-60%  40-60%  60-80%  20-40%  40-60%  40-60%
## [561] 20-40%  40-60%  20-40%  20-40%  20-40%  60-80%  60-80%  40-60%
## [569] 40-60%  40-60%  40-60%  60-80%  40-60%  40-60%  20-40%  40-60%
## [577] 40-60%  0-20%   20-40%  20-40%  60-80%  40-60%  20-40%  60-80%
## [585] 60-80%  40-60%  20-40%  60-80%  20-40%  40-60%  40-60%  60-80%
## [593] 20-40%  40-60%  60-80%  20-40%  0-20%   60-80%  40-60%  20-40%
## [601] 20-40%  40-60%  40-60%  20-40%  20-40%  60-80%  40-60%  60-80%
## [609] 20-40%  40-60%  20-40%  40-60%  20-40%  20-40%  20-40%  40-60%
## [617] 40-60%  20-40%  40-60%  40-60%  20-40%  20-40%  20-40%  60-80%
## [625] 40-60%  40-60%  40-60%  80-100% 20-40%  20-40%  60-80%  40-60%
## [633] 20-40%  40-60%  20-40%  20-40%  20-40%  40-60%  20-40%  60-80%
## [641] 20-40%  40-60%  40-60%  20-40%  40-60%  40-60%  40-60%  40-60%
## [649] 40-60%  40-60%  40-60%  40-60%  20-40%  40-60%  40-60%  80-100%
## [657] 80-100% 0-20%   20-40%  40-60%  60-80%  20-40%  40-60%  20-40%
## [665] 60-80%  60-80%  20-40%  20-40%  40-60%  40-60%  60-80%  20-40%
## [673] 60-80%  20-40%  80-100% 60-80%  20-40%  40-60%  40-60%  20-40%
## [681] 40-60%  40-60%  20-40%  60-80%  40-60%  20-40%  20-40%  20-40%
## [689] 40-60%  40-60%  20-40%  20-40%  60-80%  60-80%  40-60%  40-60%
## [697] 60-80%  40-60%  40-60%  40-60%  0-20%   20-40%  60-80%  20-40%
## [705] 40-60%  60-80%  40-60%  40-60%  40-60%  40-60%  40-60%  20-40%
## [713] 40-60%  60-80%  60-80%  40-60%  40-60%  60-80%  40-60%  40-60%
## [721] 40-60%  20-40%  40-60%  60-80%  40-60%  20-40%  40-60%  40-60%
## [729] 40-60%  60-80%  20-40%  40-60%  40-60%  0-20%   20-40%  20-40%
## [737] 20-40%  60-80%  20-40%  40-60%  40-60%  20-40%  80-100% 40-60%
## [745] 20-40%  20-40%  20-40%  20-40%  40-60%  60-80%  40-60%  60-80%
## [753] 40-60%  60-80%  40-60%  20-40%  0-20%   80-100% 20-40%
## Levels: 0-20% 20-40% 40-60% 60-80% 80-100%
```

```
allENG$section = mapvalues(allENG$section, from = c(0, 1, 2, 3, 4), to = c("0-20%",
"20-40%", "40-60%", "60-80%", "80-100%"))
table(allENG$section)
```

```
##
##   0-20%  20-40%  40-60%  60-80% 80-100%
##      33     288     313      81      20
```

```
str(allENG$section)
```

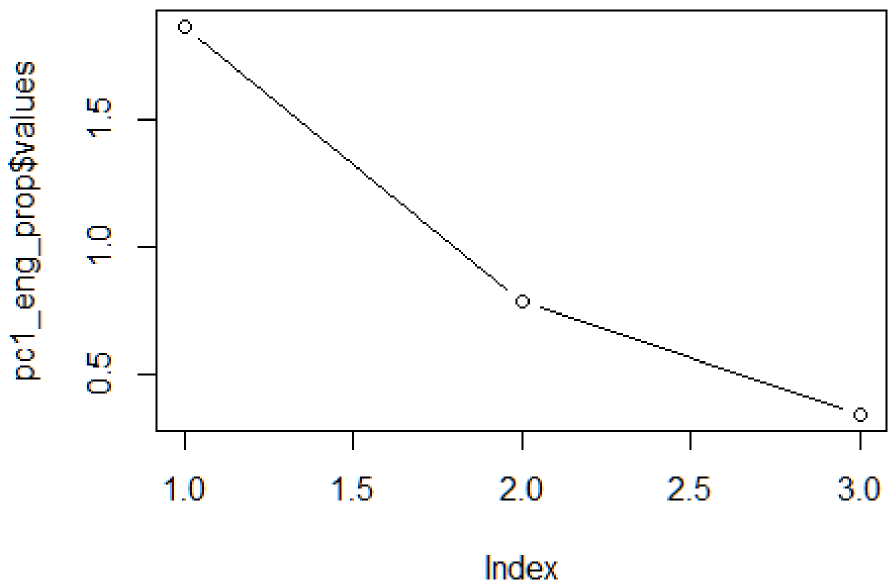
```
##  Factor w/ 5 levels "0-20%","20-40%",..: NA 3 2 3 3 3 4 3 2 3 ...
```

```
counts <- table(allENG$section, allENG$catmain)
counts

##
##           conc.a prop.a conc.h prop.h conc.v prop.v
##   0-20%        6      2     10      5      6      4
##   20-40%      60     27     61     42     64     34
##   40-60%      54     54     58     44     52     51
##   60-80%       7     23      4     18      9     20
##   80-100%      0      8      0      5      1      6

counts = prop.table(counts, 2)

# below, run first line, then return and keep running:
png(file="stacked_exc_eng.png", units="in", width=6, height=6, res=1000)
par(mar=c(2,-.3,3,-.3)+.4)  # run twice, if necessary
barplot(counts, width=10, main = 'Modality exclusivity of English properties and concepts
per dominant modality (Y axis = n)      ', legend = rownames(counts), xlim=c(0,100),
axes=FALSE, args.legend = list(x = "topright", bty = "n", inset=c(.1, .2)))
dev.off()

## png
##   2

# See in folder and compare.


# Comparison English Dutch on exclusivity
# Properties
t.test(props$exc_eng, mu = 0.40)

##
##  One Sample t-test
##
## data:  props$exc_eng
## t = 8.5326, df = 335, p-value = 5.031e-16
## alternative hypothesis: true mean is not equal to 0.4
## 95 percent confidence interval:
##  0.4626335 0.5001642
## sample estimates:
## mean of x
## 0.4813988

# The difference is considerable, t(734) = 18.8, p < .001
# dz = t/vn = 0.47

# Concepts
t.test(concs$exc_eng, mu = 0.29)

##
##  One Sample t-test
```

```
##
## data:  concs$exc_eng
## t = 16.857, df = 386, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0.29
## 95 percent confidence interval:
##  0.3831297 0.4077230
## sample estimates:
## mean of x
## 0.3954264

# The difference is considerable, t(734) = 18.8, p < .001
# dz = t/vn = 0.83
# ___________________________________________________________________________

# RELATION AMONG MODALITIES

# Below, very informative plots based on Principal Components Analysis (PCA),
# as in Lynott and Connell (2009, 2013)
# Firstly it is performed on the Dutch norms, then on the English ones, leaving out
# gustatory and olfactory scores and words. At the end, Dutch and English plots are
# compared.

all <- read.csv('all.csv')
nrow(all) # 747 used in Dutch norms + English not used

## [1] 759

# ON ENGLISH NORMS
# PCA plotting on the English norms, as based on Lynott and Connell's
# supplementary materials (http://www.lancaster.ac.uk/people/connelll/lab/norms.html).

# ENG PROPERTIES
# check conditions for a PCA
# matrix

eng_prop <- all[all$cat == 'prop', c('Aud_eng', 'Hap_eng', 'Vis_eng')]
nrow(eng_prop)

## [1] 343

eng_prop_matrix <- cor(eng_prop, use = 'complete.obs')
eng_prop_matrix

##            Aud_eng    Hap_eng    Vis_eng
## Aud_eng  1.0000000 -0.4165084 -0.6247598
## Hap_eng -0.4165084  1.0000000  0.2344421
## Vis_eng -0.6247598  0.2344421  1.0000000

round(eng_prop_matrix, 2)
```

```
##         Aud_eng Hap_eng Vis_eng
## Aud_eng    1.00   -0.42   -0.62
## Hap_eng   -0.42    1.00    0.23
## Vis_eng   -0.62    0.23    1.00

# OK: correlations good for a PCA, with enough < .3

# now on the raw vars:
nrow(eng_prop)

## [1] 343

cortest.bartlett(eng_prop)

## R was not square, finding R from data

## $chisq
## [1] 233.5851
##
## $p.value
## [1] 2.320842e-50
##
## $df
## [1] 3

# GOOD: Bartlett's test significant

# KMO: Kaiser-Meyer-Olkin Measure of Sampling Adequacy
KMO(eng_prop_matrix)

## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = eng_prop_matrix)
## Overall MSA =  0.56
## MSA for each item =
## Aud_eng Hap_eng Vis_eng
##    0.54    0.64    0.55

# Result: .56 = mediocre. PCA not strongly recommended. But we still do it
# because the purpose is graphical only.

# check determinant
det(eng_prop_matrix)

## [1] 0.5032448

# GOOD: >0.00001

# start off with unrotated PCA
pc1_eng_prop <- psych::principal(eng_prop, nfactors = 3, rotate = "none")
pc1_eng_prop
```

```
## Principal Components Analysis
## Call: psych::principal(r = eng_prop, nfactors = 3, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##           PC1   PC2  PC3 h2       u2 com
## Aud_eng -0.89  0.13 0.44  1 -2.2e-16 1.5
## Hap_eng  0.64  0.75 0.15  1  1.1e-16 2.0
## Vis_eng  0.81 -0.46 0.36  1 -4.4e-16 2.0
##
##                        PC1  PC2  PC3
## SS loadings           1.87 0.79 0.34
## Proportion Var        0.62 0.26 0.11
## Cumulative Var        0.62 0.89 1.00
## Proportion Explained  0.62 0.26 0.11
## Cumulative Proportion 0.62 0.89 1.00
##
## Mean item complexity =  1.9
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0
##  with the empirical chi square  0  with prob <  NA
##
## Fit based upon off diagonal values = 1

# RESULT: Extract either one PC, acc to Kaiser's criterion, or two RCs, acc to
# Joliffe's (Field, Miles, & Field, 2012)

# Unrotated: scree plot
plot(pc1_eng_prop$values, type = "b")
```




```
# Result: again one or two RCs should be extracted

# Now with varimax rotation, Kaiser-normalized (by default)
pc2_eng_prop <- psych::principal(eng_prop, nfactors = 2, rotate = "varimax",
scores = TRUE)
pc2_eng_prop

## Principal Components Analysis
## Call: psych::principal(r = eng_prop, nfactors = 2, rotate = "varimax",
##     scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##           RC1   RC2   h2    u2 com
## Aud_eng -0.82 -0.36 0.81 0.190 1.4
## Hap_eng  0.16  0.98 0.98 0.022 1.1
## Vis_eng  0.93  0.04 0.87 0.130 1.0
##
##                        RC1  RC2
## SS loadings           1.57 1.09
## Proportion Var        0.52 0.36
## Cumulative Var        0.52 0.89
## Proportion Explained  0.59 0.41
## Cumulative Proportion 0.59 1.00
##
## Mean item complexity =  1.1
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.1
```

```
##  with the empirical chi square  21.7  with prob <  NA
##
## Fit based upon off diagonal values = 0.95
```

```r
pc2_eng_prop$loadings
```

```
##
## Loadings:
##         RC1    RC2
## Aud_eng -0.825 -0.360
## Hap_eng  0.156  0.977
## Vis_eng  0.932
##
##                  RC1   RC2
## SS loadings    1.573 1.085
## Proportion Var 0.524 0.362
## Cumulative Var 0.524 0.886
```

```r
# two components are good, as they both have eigenvalues over 1

pc2_eng_prop$residual
```

```
##            Aud_eng    Hap_eng    Vis_eng
## Aud_eng 0.18971667 0.06403298 0.15723160
## Hap_eng 0.06403298 0.02161235 0.05306865
## Vis_eng 0.15723160 0.05306865 0.13030894
```

```r
pc2_eng_prop$fit
```

```
## [1] 0.9724565
```

```r
pc2_eng_prop$communality
```

```
##   Aud_eng   Hap_eng   Vis_eng
## 0.8102833 0.9783877 0.8696911
```

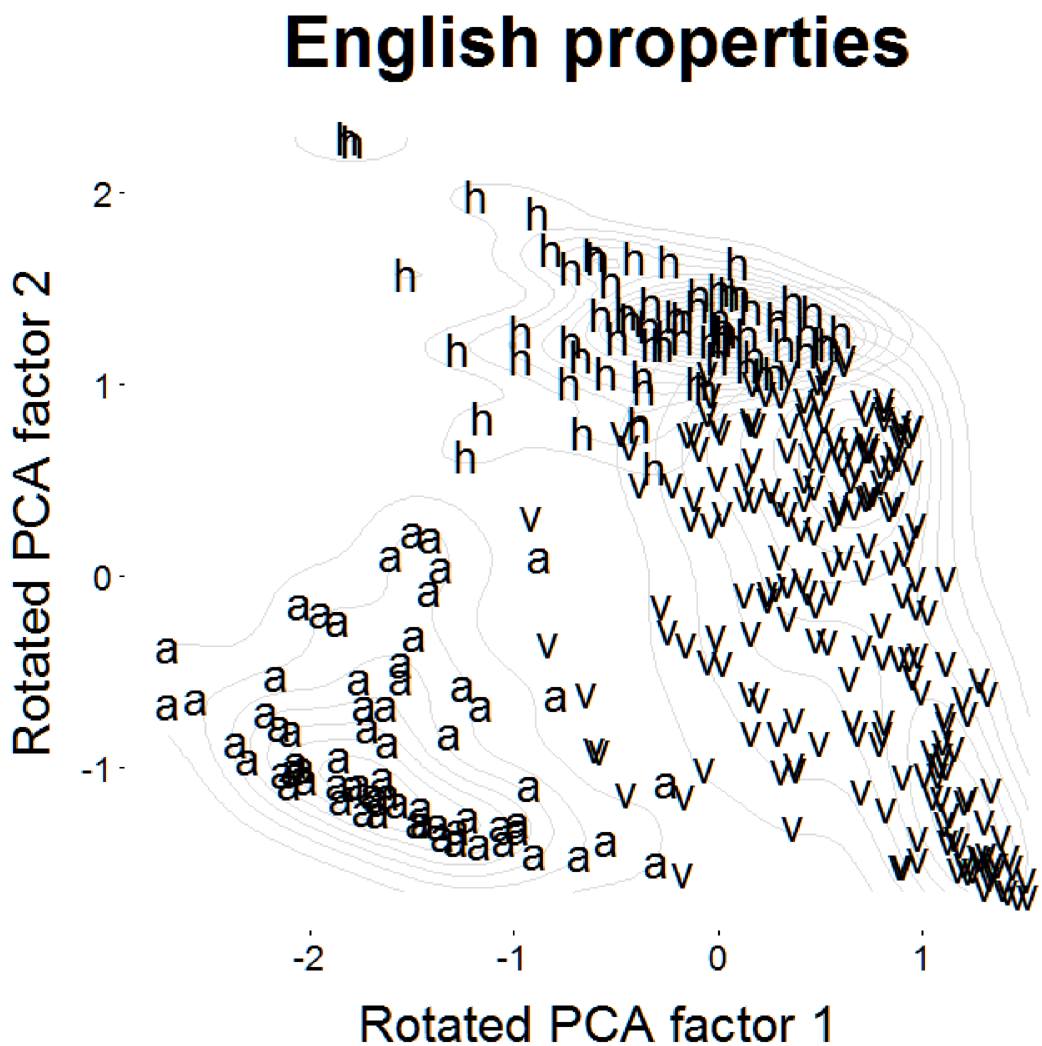
```r
# Results based on a Kaiser-normalizalized orthogonal (varimax) rotation
# (by default in psych::stats). Residuals bad: more than 50% have absolute
# values > 0.05. Model fit good, > .90. Communalities good,
# all > .7.

# subset and add PCs
eng_props <- all[all$cat == 'prop', ]
nrow(eng_props)
```

```
## [1] 343
```

```r
eng_props <- cbind(eng_props, pc2_eng_prop$scores)
nrow(eng_props)
```

```
## [1] 343
```

```r
# Finally, plot
Engprops <- ggplot(eng_props,
  aes(RC1, RC2, label = as.character(main_eng))) +
  aes (x = RC1, y = RC2, by = main_eng) + stat_density2d (color = "gray87") +
  geom_text(size = 7) +
    ggtitle ('English properties') +
    theme_bw() +    # theme with white background
    theme(    # clear background, gridlines, chart border
    plot.background = element_blank()
   ,panel.grid.major = element_blank()
   ,panel.grid.minor = element_blank()
   ,panel.border = element_blank()
  ) +
  theme(axis.line = element_line(color = 'black')) +  # draw x and y lines
    theme(axis.title.x = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.title.y = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.text.x  = element_text(size=16),
       axis.text.y  = element_text(size=16)) +
  labs(x = "Rotated PCA factor 1", y = "Rotated PCA factor 2") +
    theme(plot.title = element_text(size = 32, face = "bold",
    margin=margin(15,15,15,15)))

plot(Engprops)  # ! THE PLOT IS SHOWN BADLY ON HERE. PLEASE SEE THE SAVED PLOT + THEN
```



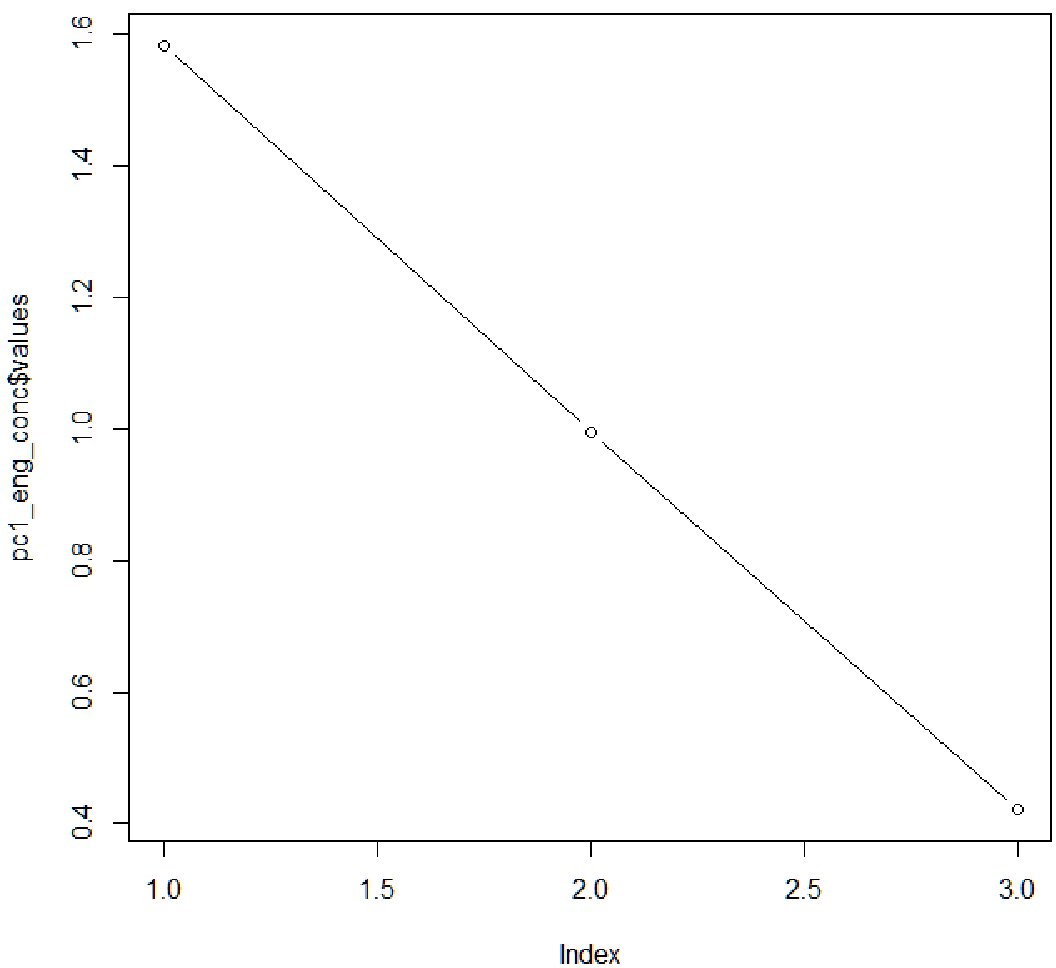

```
# THE COMBINED PLOTS

# Now to save, run first line below and return to keep running. See your folder.
png(file="Engprops_highres.png", units="in", width=13, height=13, res=900)
plot(Engprops)
dev.off()

## png
##   2

# Adjust for combined plots:

Engprops4 <- ggplot(eng_props,
  aes(RC1, RC2, label = as.character(main_eng)) +
  aes (x = RC1, y = RC2, by = main_eng) + stat_density2d (color = "gray87") +
  geom_text(size = 7) +
    ggtitle ('English properties') +
```

```
    theme_bw() +    # theme with white background
    theme(    # clear background, gridlines, chart border
    plot.background = element_blank()
   ,panel.grid.major = element_blank()
   ,panel.grid.minor = element_blank()
   ,panel.border = element_blank()
  ) +
  theme(axis.line = element_line(color = 'black')) +  # draw x and y lines
    theme(axis.title.x = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
          axis.title.y = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
          axis.text.x  = element_text(size=16),
        axis.text.y  = element_text(size=16)) +
  labs(x = "", y = "Rotated PCA factor 2") +
    theme(plot.title = element_text(size = 32, face = "bold",
    margin=margin(15,15,15,15)))


# ENG CONCEPTS
# check conditions for a PCA
# matrix
eng_conc <- all[all$cat == 'conc', c('Aud_eng', 'Hap_eng', 'Vis_eng')]
nrow(eng_conc)

## [1] 416

eng_conc_matrix <- cor(eng_conc, use = 'complete.obs')
eng_conc_matrix

##              Aud_eng    Hap_eng      Vis_eng
## Aud_eng  1.000000000 -0.1760092 -0.008395838
## Hap_eng -0.176009214  1.0000000  0.554494445
## Vis_eng -0.008395838  0.5544944  1.000000000

round(eng_conc_matrix, 2)

##         Aud_eng Hap_eng Vis_eng
## Aud_eng    1.00   -0.18   -0.01
## Hap_eng   -0.18    1.00    0.55
## Vis_eng   -0.01    0.55    1.00

# POOR: correlations not apt for a PCA, with too many below .3

# now on the raw data:
nrow(eng_conc)

## [1] 416

cortest.bartlett(eng_conc)

## R was not square, finding R from data
```

```
## $chisq
## [1] 169.7255
##
## $p.value
## [1] 1.458581e-36
##
## $df
## [1] 3

# GOOD: Bartlett's test significant

# KMO: Kaiser-Meyer-Olkin Measure of Sampling Adequacy
KMO(eng_conc_matrix)

## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = eng_conc_matrix)
## Overall MSA =  0.48
## MSA for each item =
## Aud_eng Hap_eng Vis_eng
##    0.36    0.49    0.48

# Result: .48 = poor. PCA not strongly recommended. But we still do it
# because the purpose is graphical really.

# check determinant
det(eng_conc_matrix)

## [1] 0.663125

# GOOD: >0.00001

# start off with unrotated PCA
pc1_eng_conc <- psych::principal(eng_conc, nfactors = 3, rotate = "none")
pc1_eng_conc

## Principal Components Analysis
## Call: psych::principal(r = eng_conc, nfactors = 3, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##           PC1  PC2   PC3 h2       u2 com
## Aud_eng -0.28 0.95  0.13  1  0.0e+00 1.2
## Hap_eng  0.89 0.01  0.46  1 -2.2e-16 1.5
## Vis_eng  0.85 0.30 -0.44  1  0.0e+00 1.8
##
##                        PC1  PC2  PC3
## SS loadings           1.58 1.00 0.42
## Proportion Var        0.53 0.33 0.14
## Cumulative Var        0.53 0.86 1.00
## Proportion Explained  0.53 0.33 0.14
## Cumulative Proportion 0.53 0.86 1.00
##
## Mean item complexity =  1.5
```

```
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0
##  with the empirical chi square  0  with prob <   NA
##
## Fit based upon off diagonal values = 1
```

```r
# RESULT: Extract either one PC, acc to Kaiser's criterion, or two RCs, acc to
# Joliffe's (Field, Miles, & Field, 2012)

# Unrotated: scree plot
plot(pc1_eng_conc$values, type = "b")
```

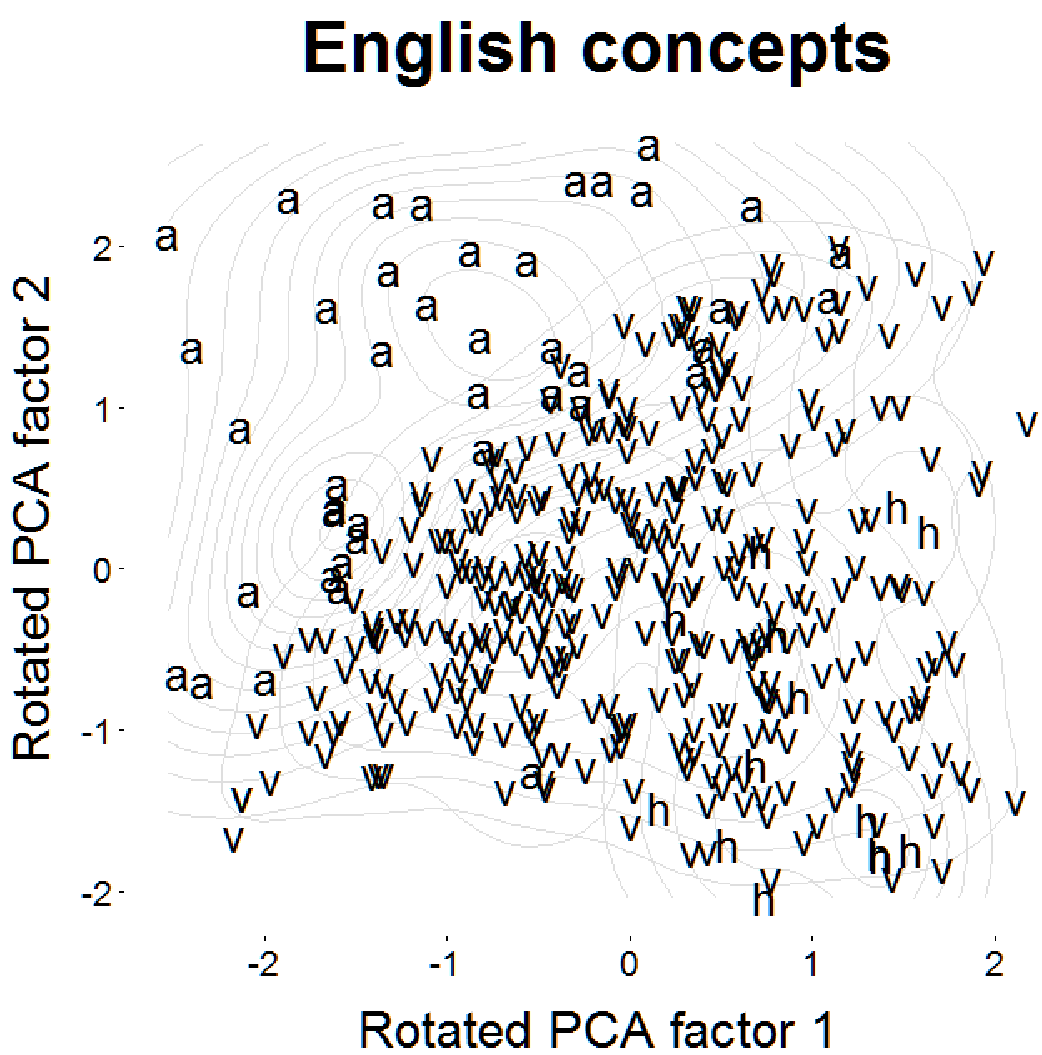



```r
# Result: two PCs obtain.

# Now with varimax rotation, Kaiser-normalized (by default):
# always preferable because it captures explained variance best.
```

```r
pc2_eng_conc <- psych::principal(eng_conc, nfactors = 2, rotate = "varimax",
scores = TRUE)
pc2_eng_conc

## Principal Components Analysis
## Call: psych::principal(r = eng_conc, nfactors = 2, rotate = "varimax",
##     scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##           RC1   RC2   h2    u2 com
## Aud_eng -0.04  0.99 0.98 0.018 1.0
## Hap_eng  0.87 -0.20 0.79 0.211 1.1
## Vis_eng  0.89  0.09 0.81 0.192 1.0
##
##                        RC1  RC2
## SS loadings           1.55 1.03
## Proportion Var        0.52 0.34
## Cumulative Var        0.52 0.86
## Proportion Explained  0.60 0.40
## Cumulative Proportion 0.60 1.00
##
## Mean item complexity =  1
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.13
##  with the empirical chi square  39.63  with prob <  NA
##
## Fit based upon off diagonal values = 0.86

pc2_eng_conc$loadings

##
## Loadings:
##         RC1    RC2
## Aud_eng         0.990
## Hap_eng  0.865 -0.201
## Vis_eng  0.894
##
##                  RC1   RC2
## SS loadings    1.551 1.029
## Proportion Var 0.517 0.343
## Cumulative Var 0.517 0.860

pc2_eng_conc$residual

##             Aud_eng     Hap_eng     Vis_eng
## Aud_eng  0.01775551  0.06123023 -0.05834261
## Hap_eng  0.06123023  0.21115367 -0.20119566
## Vis_eng -0.05834261 -0.20119566  0.19170728

pc2_eng_conc$fit
```

```
## [1] 0.9518855

pc2_eng_conc$communality

##   Aud_eng   Hap_eng   Vis_eng
## 0.9822445 0.7888463 0.8082927

# Results based on a Kaiser-normalizalized orthogonal (varimax) rotation
# (by default in psych::stats). Residuals bad: over 50% have absolute
# values > 0.05. Model fit good, > .90. Communalities good, all > .7.

# subset and add PCs
eng_concs <- all[all$cat == 'conc', ]
nrow(eng_concs)

## [1] 416

eng_concs <- cbind(eng_concs, pc2_eng_conc$scores)
summary(eng_concs$RC1, eng_concs$RC2)

##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
## -2.52800 -0.72020  0.03826  0.00000  0.72380  2.18400       24

eng_concs <- eng_concs[eng_concs$normed == 'Dut_Eng' | eng_concs$normed ==
'English',]
nrow(eng_concs)

## [1] 392

summary(eng_concs$RC1, eng_concs$RC2)

##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.52800 -0.72020  0.03826  0.00000  0.72380  2.18400

# Finally, plot
Engconcs <- ggplot(eng_concs,
  aes(RC1, RC2, label = as.character(main_eng))) +
  aes (x = RC1, y = RC2, by = main_eng) + stat_density2d (color = "gray87") +
  geom_text(size = 7) +
    ggtitle ('English concepts') +
    theme_bw() +    # theme with white background
    theme(    # clear background, gridlines, chart border
    plot.background = element_blank()
   ,panel.grid.major = element_blank()
   ,panel.grid.minor = element_blank()
   ,panel.border = element_blank()
  ) +
  theme(axis.line = element_line(color = 'black')) +  # draw x and y lines
    theme(axis.title.x = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.title.y = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
```

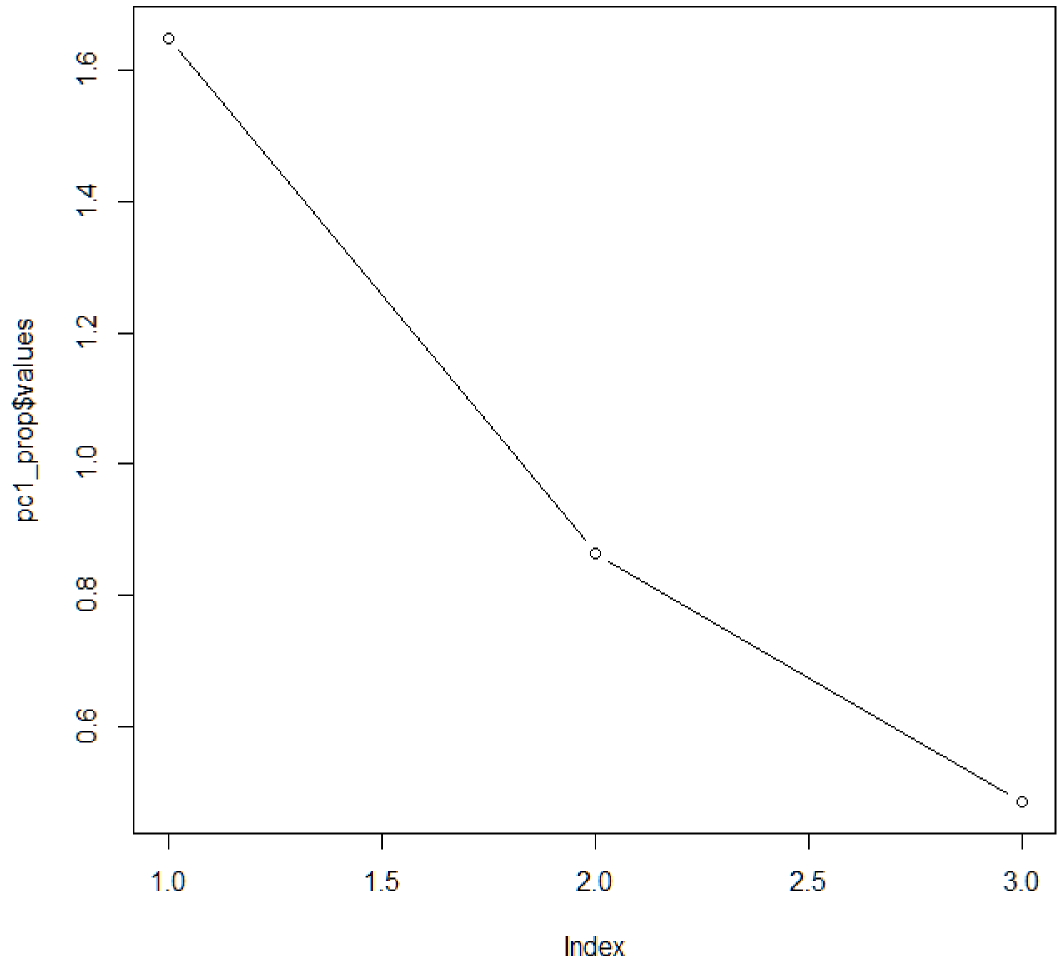
```
        axis.text.x  = element_text(size=16),
      axis.text.y  = element_text(size=16)) +
  labs(x = "Rotated PCA factor 1", y = "Rotated PCA factor 2") +
    theme(plot.title = element_text(size = 32, face = "bold",
    margin=margin(15,15,15,15)))

Engconcs  # ! THE PLOT IS SHOWN BADLY ON HERE. PLEASE SEE THE SAVED PLOT
```




```
# Now to save, run first line below and return to keep running. See your folder.
png(file="Engconcs_highres.png", units="in", width=13, height=13, res=900)
plot(Engconcs)
dev.off()

## png
##   2
```

```r
# ON DUTCH NORMS

# properties
# check conditions for a PCA

# matrix
prop <- all[all$cat == 'prop', c('Auditory', 'Haptic', 'Visual')]
nrow(prop)

## [1] 343

prop_matrix <- cor(prop, use = 'complete.obs')
prop_matrix

##             Auditory     Haptic     Visual
## Auditory  1.0000000 -0.2280165 -0.5134304
## Haptic   -0.2280165  1.0000000  0.1930402
## Visual   -0.5134304  0.1930402  1.0000000

round(prop_matrix, 2)

##          Auditory Haptic Visual
## Auditory     1.00  -0.23  -0.51
## Haptic      -0.23   1.00   0.19
## Visual      -0.51   0.19   1.00

# POOR: correlations not apt for a PCA, with too many below .3

# now on the raw vars:
nrow(prop)

## [1] 343

cortest.bartlett(prop)

## R was not square, finding R from data

## $chisq
## [1] 125.0759
##
## $p.value
## [1] 6.224181e-27
##
## $df
## [1] 3

# GOOD: Bartlett's test significant

# KMO: Kaiser-Meyer-Olkin Measure of Sampling Adequacy
KMO(prop_matrix)
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = prop_matrix)
## Overall MSA =  0.56
## MSA for each item =
## Auditory   Haptic   Visual
##     0.54     0.74     0.55

# Result: .56 = mediocre. PCA not strongly recommended. But we still do it
# because the purpose is graphical only.

# check determinant
det(prop_matrix)

## [1] 0.6923318

# GOOD: >0.00001

# start off with unrotated PCA
pc1_prop <- psych::principal(prop, nfactors = 3, rotate = "none")
pc1_prop

## Principal Components Analysis
## Call: psych::principal(r = prop, nfactors = 3, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##            PC1   PC2  PC3 h2       u2 com
## Auditory -0.83  0.23 0.50  1  0.0e+00 1.8
## Haptic    0.54  0.84 0.04  1 -1.6e-15 1.7
## Visual    0.82 -0.31 0.48  1 -8.9e-16 2.0
##
##                        PC1  PC2  PC3
## SS loadings           1.65 0.86 0.49
## Proportion Var        0.55 0.29 0.16
## Cumulative Var        0.55 0.84 1.00
## Proportion Explained  0.55 0.29 0.16
## Cumulative Proportion 0.55 0.84 1.00
##
## Mean item complexity =  1.8
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0
##  with the empirical chi square  0  with prob <  NA
##
## Fit based upon off diagonal values = 1

# RESULT: Only PC1, with eigenvalue > 1, should be extracted,
# acc to Kaiser's criterion (Jolliffe's threshold of 0.7 way too lax;
# Field, Miles, & Field, 2012)

# Unrotated: scree plot
plot(pc1_prop$values, type = "b")
```




```
# Result: one or two RCs should be extracted, converging with eigenvalues

# Now with varimax rotation, Kaiser-normalized (by default).
# Always preferable because it captures explained variance best.
# Compare eigenvalues w/ 1 & 2 factors

pc2_prop <- psych::principal(prop, nfactors = 2, rotate = "varimax", scores = TRUE)
pc2_prop

## Principal Components Analysis
## Call: psych::principal(r = prop, nfactors = 2, rotate = "varimax",
##     scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##            RC1   RC2   h2     u2 com
## Auditory -0.85 -0.16 0.75 0.2498 1.1
## Haptic    0.11  0.99 1.00 0.0016 1.0
## Visual    0.87  0.08 0.77 0.2337 1.0
```

```
##
##                        RC1  RC2
## SS loadings           1.5 1.02
## Proportion Var        0.5 0.34
## Cumulative Var        0.5 0.84
## Proportion Explained  0.6 0.40
## Cumulative Proportion 0.6 1.00
##
## Mean item complexity =  1
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.14
##  with the empirical chi square  40.59  with prob <  NA
##
## Fit based upon off diagonal values = 0.83
```

```
pc2_prop$loadings
```

```
##
## Loadings:
##          RC1    RC2
## Auditory -0.852 -0.158
## Haptic    0.107  0.993
## Visual    0.872
##
##                  RC1   RC2
## SS loadings    1.497 1.018
## Proportion Var 0.499 0.339
## Cumulative Var 0.499 0.838
```

```
# good to extract 2 factors, as they both explain quite the same variance,
# and both surpass 1 eigenvalue


pc2_prop$residual
```

```
##             Auditory      Haptic     Visual
## Auditory 0.24984223 0.020051207 0.24163179
## Haptic   0.02005121 0.001609219 0.01939227
## Visual   0.24163179 0.019392274 0.23369117
```

```
pc2_prop$fit
```

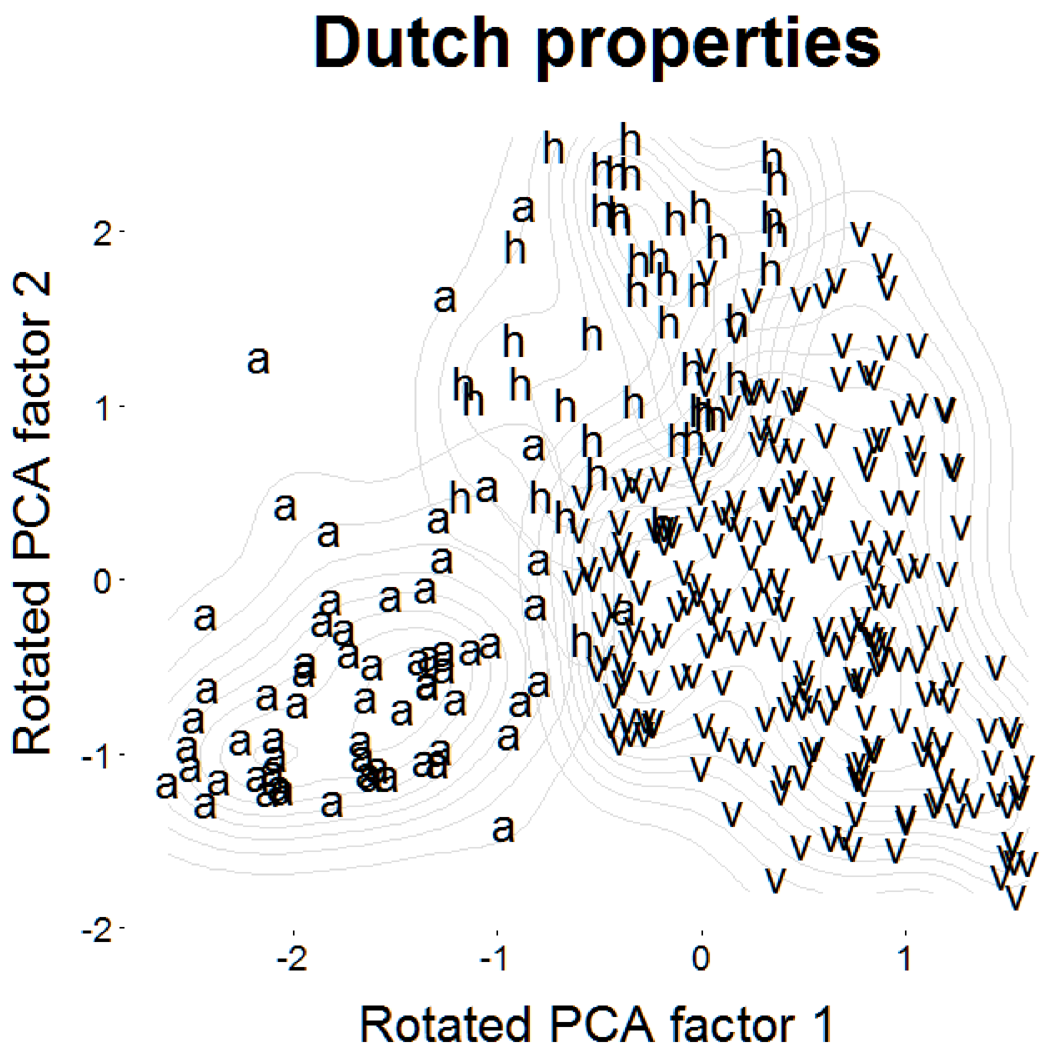
```
## [1] 0.9364867
```

```
pc2_prop$communality
```

```
##  Auditory    Haptic    Visual
## 0.7501578 0.9983908 0.7663088
```

```
# Results based on a Kaiser-normalizalized orthogonal (varimax) rotation
# (by default in psych::stats). Residuals OK: fewer than 50% have absolute
```

```
# values > 0.05 (exactly 50% do).Model fit good, > .90.
# Communalities good, all > .7 (av = .83).

# subset and add PCs
props <- all[all$cat == 'prop', ]
nrow(props)

## [1] 343

props <- cbind(props, pc2_prop$scores)
nrow(props)

## [1] 343

# Finally, plot: letters+density (cf. Lynott & Connell, 2009, 2013)

NLprops <- ggplot(props,
  aes(RC1, RC2, label = as.character(main))) +
  aes (x = RC1, y = RC2, by = main) + stat_density2d (color = "gray87") +
  geom_text(size = 7) +
    ggtitle ('Dutch properties') +
    theme_bw() +    # theme with white background
    theme(    # clear background, gridlines, chart border
    plot.background = element_blank()
   ,panel.grid.major = element_blank()
   ,panel.grid.minor = element_blank()
   ,panel.border = element_blank()
  ) +
  theme(axis.line = element_line(color = 'black')) +  # draw x and y lines
    theme(axis.title.x = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.title.y = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.text.x  = element_text(size=16),
       axis.text.y  = element_text(size=16)) +
  labs(x = "Rotated PCA factor 1", y = "Rotated PCA factor 2") +
    theme(plot.title = element_text(size = 32, face = "bold",
    margin=margin(15,15,15,15)))

NLprops  # ! THE PLOT IS SHOWN BADLY ON HERE. PLEASE SEE THE SAVED PLOT

## Warning: Removed 7 rows containing non-finite values (stat_density2d).

## Warning: Removed 7 rows containing missing values (geom_text).
```

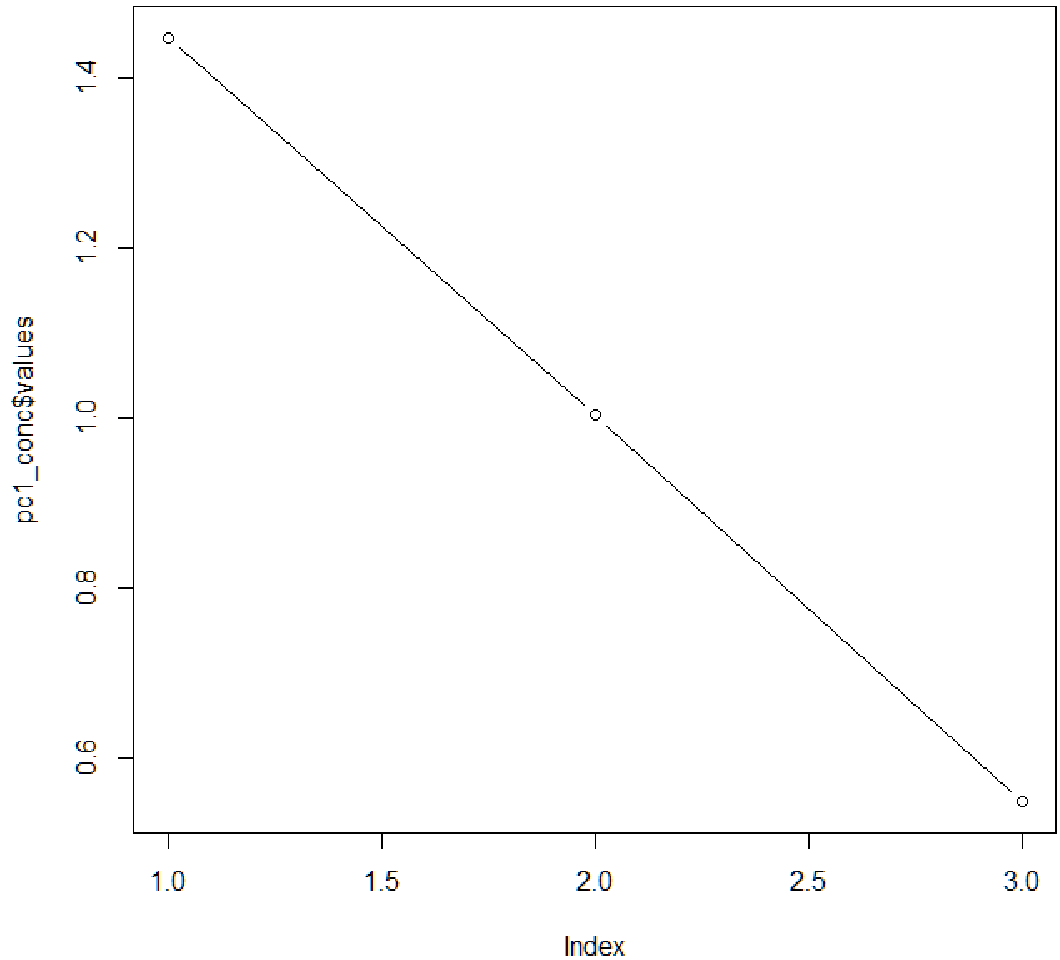



```
#  Now to save, run first line below and return to keep running. See your folder.
png(file="NLprops_highres.png", units="in", width=13, height=13, res=900)
plot(NLprops)

## Warning: Removed 7 rows containing non-finite values (stat_density2d).

## Warning: Removed 7 rows containing missing values (geom_text).

# warning normal: just removing English properties not used in Dutch
dev.off()

## png
##   2

# Adjust for combined plots:

NLprops2 <- ggplot(props,
```

```
  aes(RC1, RC2, label = as.character(main))) +
  aes (x = RC1, y = RC2, by = main) + stat_density2d (color = "gray87") +
  geom_text(size = 7) +
    ggtitle ('Dutch properties') +
    theme_bw() +    # theme with white background
    theme(    # clear background, gridlines, chart border
    plot.background = element_blank()
   ,panel.grid.major = element_blank()
   ,panel.grid.minor = element_blank()
   ,panel.border = element_blank()
  ) +
  theme(axis.line = element_line(color = 'black')) +  # draw x and y lines
    theme(axis.title.x = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.title.y = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.text.x  = element_text(size=16),
       axis.text.y  = element_text(size=16)) +
  labs(x = "Rotated PCA factor 1", y = "") +
    theme(plot.title = element_text(size = 32, face = "bold",
    margin=margin(15,15,15,15)))

# Next:

NLprops4 <- ggplot(props,
  aes(RC1, RC2, label = as.character(main))) +
  aes (x = RC1, y = RC2, by = main) + stat_density2d (color = "gray87") +
  geom_text(size = 7) +
    ggtitle ('Dutch properties') +
    theme_bw() +    # theme with white background
    theme(    # clear background, gridlines, chart border
    plot.background = element_blank()
   ,panel.grid.major = element_blank()
   ,panel.grid.minor = element_blank()
   ,panel.border = element_blank()
  ) +
  theme(axis.line = element_line(color = 'black')) +  # draw x and y lines
    theme(axis.title.x = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.title.y = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.text.x  = element_text(size=16),
       axis.text.y  = element_text(size=16)) +
  labs(x = "", y = "") +
    theme(plot.title = element_text(size = 32, face = "bold",
    margin=margin(15,15,15,15)))
```

```r
# CONCEPTS
# check conditions for a PCA
# matrix
conc <- all[all$cat == 'conc', c('Auditory', 'Haptic', 'Visual')]
nrow(conc)

## [1] 416

conc_matrix <- cor(conc, use = 'complete.obs')
conc_matrix

##              Auditory       Haptic     Visual
## Auditory  1.000000000 -0.008508063 0.08486561
## Haptic   -0.008508063  1.000000000 0.44144835
## Visual    0.084865608  0.441448353 1.00000000

round(conc_matrix, 2)

##          Auditory Haptic Visual
## Auditory     1.00  -0.01   0.08
## Haptic      -0.01   1.00   0.44
## Visual       0.08   0.44   1.00

# POOR: correlations not apt for a PCA, with too many below .3

# now on the raw data:
nrow(conc)

## [1] 416

cortest.bartlett(conc)

## R was not square, finding R from data

## $chisq
## [1] 93.63824
##
## $p.value
## [1] 3.621992e-20
##
## $df
## [1] 3

# GOOD: Bartlett's test significant

# KMO: Kaiser-Meyer-Olkin Measure of Sampling Adequacy
KMO(conc_matrix)

## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = conc_matrix)
## Overall MSA =  0.49
## MSA for each item =
```

```
## Auditory   Haptic   Visual
##     0.37     0.49     0.49

# Result: .49 = poor. PCA not strongly recommended. But we still do it
# because the purpose is graphical really.

# check determinant
det(conc_matrix)

## [1] 0.7972113

# GOOD: >0.00001

# start off with unrotated PCA
pc1_conc <- psych::principal(conc, nfactors = 3, rotate = "none")
pc1_conc

## Principal Components Analysis
## Call: psych::principal(r = conc, nfactors = 3, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##           PC1   PC2   PC3 h2      u2 com
## Auditory 0.15  0.98  0.11  1 1.1e-16 1.1
## Haptic   0.84 -0.19  0.51  1 4.4e-16 1.8
## Visual   0.85  0.02 -0.52  1 0.0e+00 1.7
##
##                        PC1  PC2  PC3
## SS loadings           1.45 1.00 0.55
## Proportion Var        0.48 0.33 0.18
## Cumulative Var        0.48 0.82 1.00
## Proportion Explained  0.48 0.33 0.18
## Cumulative Proportion 0.48 0.82 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0
##  with the empirical chi square  0  with prob <  NA
##
## Fit based upon off diagonal values = 1

# RESULT good: PC1 and PC2, with eigenvalue > 1, should be extracted,
# acc to Kaiser's criterion (Jolliffe's threshold of 0.7 way too lax;
# Field, Miles, & Field, 2012)

# Unrotated: scree plot
plot(pc1_conc$values, type = "b")
```




```
# Result: with no point of inflexion along the y axis, two PCs would obtain.

# Now with varimax rotation, Kaiser-normalized (by default):
# Always preferable because it captures explained variance best.
# Compare eigenvalues w/ 1 & 2 factors

pc2_conc <- psych::principal(conc, nfactors = 2, rotate = "varimax", scores = TRUE)
pc2_conc

## Principal Components Analysis
## Call: psych::principal(r = conc, nfactors = 2, rotate = "varimax",
##     scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##            RC1   RC2   h2    u2 com
## Auditory 0.03  0.99 0.99 0.012   1
## Haptic   0.85 -0.09 0.74 0.264   1
## Visual   0.84  0.12 0.73 0.273   1
```

```
##
##                          RC1  RC2
## SS loadings           1.44 1.01
## Proportion Var        0.48 0.34
## Cumulative Var        0.48 0.82
## Proportion Explained  0.59 0.41
## Cumulative Proportion 0.59 1.00
##
## Mean item complexity =  1
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.16
##  with the empirical chi square  65.21  with prob <  NA
##
## Fit based upon off diagonal values = 0.61
```

```
pc2_conc$loadings
```

```
##
## Loadings:
##          RC1    RC2
## Auditory         0.994
## Haptic    0.854
## Visual    0.844  0.120
##
##                  RC1   RC2
## SS loadings    1.442 1.010
## Proportion Var 0.481 0.337
## Cumulative Var 0.481 0.817
```

```
# good to extract 2 factors, as they both explain quite the same variance,
# and both surpass 1 eigenvalue

pc2_conc$residual
```

```
##             Auditory     Haptic      Visual
## Auditory  0.01167046  0.0554842 -0.05648251
## Haptic    0.05548420  0.2637854 -0.26853166
## Visual   -0.05648251 -0.2685317  0.27336330
```

```
pc2_conc$fit
```

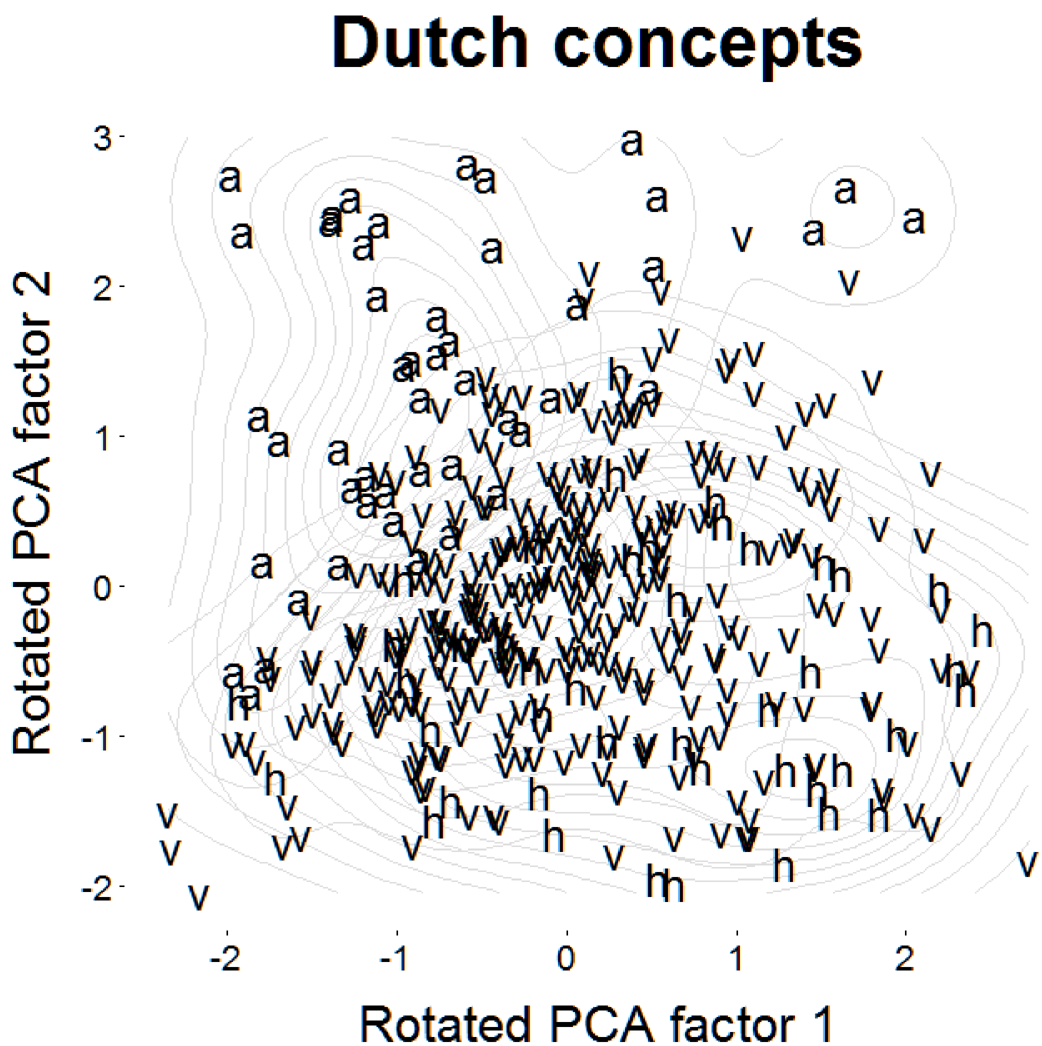
```
## [1] 0.911523
```

```
pc2_conc$communality
```

```
##  Auditory    Haptic    Visual
## 0.9883295 0.7362146 0.7266367
```

```
# Results based on a Kaiser-normalizalized orthogonal (varimax) rotation
# (by default in psych::stats). Residuals bad: over 50% have absolute
# values > 0.05. Model fit good, > .90. Communalities good, all > .7 (av = .82).
```

```
# subset and add PCs
concs <- all[all$cat == 'conc', ]
nrow(concs)

## [1] 416

concs <- cbind(concs, pc2_conc$scores)
nrow(concs)

## [1] 416

# Finally, plot
NLconcs <- ggplot(concs,
  aes(RC1, RC2, label = as.character(main))) +
  aes (x = RC1, y = RC2, by = main) + stat_density2d (color = "gray87") +
  geom_text(size = 7) +
    ggtitle ('Dutch concepts') +
    theme_bw() +    # theme with white background
    theme(    # clear background, gridlines, chart border
    plot.background = element_blank()
   ,panel.grid.major = element_blank()
   ,panel.grid.minor = element_blank()
   ,panel.border = element_blank()
  ) +
  theme(axis.line = element_line(color = 'black')) +  # draw x and y lines
    theme(axis.title.x = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.title.y = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.text.x  = element_text(size=16),
       axis.text.y  = element_text(size=16)) +
  labs(x = "Rotated PCA factor 1", y = "Rotated PCA factor 2") +
    theme(plot.title = element_text(size = 32, face = "bold",
    margin=margin(15,15,15,15)))

NLconcs  # ! THE PLOT IS SHOWN BADLY ON HERE. PLEASE SEE THE SAVED PLOT

## Warning: Removed 5 rows containing non-finite values (stat_density2d).

## Warning: Removed 5 rows containing missing values (geom_text).
```




```
#  Now to save, run first line below and return to keep running. See your folder.
png(file="NLconcs_highres.png", units="in", width=13, height=13, res=900)
plot(NLconcs)

## Warning: Removed 5 rows containing non-finite values (stat_density2d).

## Warning: Removed 5 rows containing missing values (geom_text).

# warning normal: just removing English concepts not used in Dutch
dev.off()

## png
##   2

# Adjust for combined plots:

NLconcs2 <- ggplot(concs,
```

```
  aes(RC1, RC2, label = as.character(main))) +
  aes (x = RC1, y = RC2, by = main) + stat_density2d (color = "gray87") +
  geom_text(size = 7) +
    ggtitle ('Dutch concepts') +
    theme_bw() +    # theme with white background
    theme(    # clear background, gridlines, chart border
    plot.background = element_blank()
   ,panel.grid.major = element_blank()
   ,panel.grid.minor = element_blank()
   ,panel.border = element_blank()
  ) +
  theme(axis.line = element_line(color = 'black')) +  # draw x and y lines
    theme(axis.title.x = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.title.y = element_text(colour = 'black', size = 23,
    margin=margin(15,15,15,15)),
         axis.text.x  = element_text(size=16),
       axis.text.y  = element_text(size=16)) +
  labs(x = "Rotated PCA factor 1", y = "") +
    theme(plot.title = element_text(size = 32, face = "bold",
    margin=margin(15,15,15,15)))



# Combined plots:

# Below, run first line, get back and run next.
# High resolution (may be changed at 'res='). Beware of high memory usage.

png(file="allfour_highres.png", units="in", width=19, height=19, res=1200)
multiplot(Engprops4, Engconcs, NLprops4, NLconcs2, cols = 2)

## Warning: Removed 7 rows containing non-finite values (stat_density2d).

## Warning: Removed 7 rows containing missing values (geom_text).

## Warning: Removed 5 rows containing non-finite values (stat_density2d).

## Warning: Removed 5 rows containing missing values (geom_text).

# warning normal: just those English items that were not used in Dutch
dev.off()

## png
##   2

png(file="proppair_highres.png", units="in", width=18, height=9, res=1000)
multiplot(Engprops, NLprops2, cols = 2)

## Warning: Removed 7 rows containing non-finite values (stat_density2d).
```

```
## Warning: Removed 7 rows containing missing values (geom_text).

# warning normal: just those English items that were not used in Dutch
dev.off()

## png
##   2

png(file="concpair_highres.png", units="in", width=18, height=9, res=1000)
multiplot(Engconcs, NLconcs2, cols = 2)

## Warning: Removed 5 rows containing non-finite values (stat_density2d).

## Warning: Removed 5 rows containing missing values (geom_text).

# warning normal: just those English items that were not used in Dutch
dev.off()

## png
##   2

# Find all plots in your working directory

# With a naked eye, one can see the different relationships. The significance of
# these comparisons is notable. First, it demonstrates visually the difference
# between modality exclusivity and each of the modality strengths (which of course
# is only natural considering how modality exclusivity was calculated). The two
# variables then must be different indeed because in the exclusivity analysis, the
# visual and the auditory modalities were the most similar ones, with their higher
# exclusivities. In contrast, in the independent strengths analysis, the visual and
# the haptic modalities show a clear interlock, which leaves the auditory experience
# rather on its own.
# ______________________________________________________________________________________



# ICONICITY

# Last tests: iconicity/sound symbolism on concepts and properties separately.
# Regressions include same lexical vars (DVs) as Lynott and Connell, plus
# concreteness and age of acquisition.

# Note that the selection is based on p-value thresholds, as in L&C, but also on
# AIC, which is a bayesian, relative method more appropriate with such a large
# sample. Importantly, AIC and F/p-value criteria resulted in the same inclusions
# and exclusions for every regression.

# For both props and concs, we start with PCA with all lexical variables in order
# to isolate them, because they are intercorrelated (see Table 5 in Lynott & Connell,
# 2013)
```

```r
all <- read.csv('all.csv')
nrow(all)

## [1] 759

# Length is 759 but only 747 are from these norms. Rest are from Lynott and Connell
# (2009, 2013) for comparative analyses. These extra items do not have an id number
# in the file.

# Iconicity within properties alone, as in Lynott and Connell (2013). As a novelty,
# the iconicity analysis is hereby performed also on the Dutch properties, in
# addition to the concepts.

props <- subset(all, subset = cat == 'prop')
nrow(props)

## [1] 343

# There aren't lexical data for every single word.
# Nr of properties per lexical variable (from the Dutch items only of course)
describe(complete.cases(props[complete.cases(props$Exclusivity),]
$phonemes_DUTCHPOND))

## complete.cases(props[complete.cases(props$Exclusivity), ]$phonemes_DUTCHPOND)
##        n missing  unique
##      336       0       2
##
## FALSE (151, 45%), TRUE (185, 55%)

describe(complete.cases(props[complete.cases(props$Exclusivity),]
$phon_neighbours_DUTCHPOND))

## complete.cases(props[complete.cases(props$Exclusivity), ]$phon_neighbours_DUTCHPOND)
##        n missing  unique   value
##      336       0       1    TRUE

describe(complete.cases(props[complete.cases(props$Exclusivity),]
$orth_neighbours_DUTCHPOND))

## complete.cases(props[complete.cases(props$Exclusivity), ]$orth_neighbours_DUTCHPOND)
##        n missing  unique   value
##      336       0       1    TRUE

describe(complete.cases(props[complete.cases(props$Exclusivity),]
$freq_lg10CD_SUBTLEXNL))

## complete.cases(props[complete.cases(props$Exclusivity), ]$freq_lg10CD_SUBTLEXNL)
##        n missing  unique
##      336       0       2
##
## FALSE (46, 14%), TRUE (290, 86%)
```

```r
describe(complete.cases(props[complete.cases(props$Exclusivity),]
$freq_lg10WF_SUBTLEXNL))

## complete.cases(props[complete.cases(props$Exclusivity), ]$freq_lg10WF_SUBTLEXNL)
##        n missing  unique
##      336       0       2
##
## FALSE (46, 14%), TRUE (290, 86%)

describe(complete.cases(props[complete.cases(props$Exclusivity),]
$freq_CELEX_lem))

## complete.cases(props[complete.cases(props$Exclusivity), ]$freq_CELEX_lem)
##        n missing  unique
##      336       0       2
##
## FALSE (89, 26%), TRUE (247, 74%)

describe(complete.cases(props[complete.cases(props$Exclusivity),]
$AoA_Brysbaertetal2014))

## complete.cases(props[complete.cases(props$Exclusivity), ]$AoA_Brysbaertetal2014)
##        n missing  unique
##      336       0       2
##
## FALSE (103, 31%), TRUE (233, 69%)

describe(complete.cases(props[complete.cases(props$Exclusivity),]
$concrete_Brysbaertetal2014))

## complete.cases(props[complete.cases(props$Exclusivity), ]$concrete_Brysbaertetal2014)
##        n missing  unique
##      336       0       2
##
## FALSE (103, 31%), TRUE (233, 69%)

# M, SD
stat.desc(props$letters)

##      nbr.val     nbr.null       nbr.na          min          max
##  336.0000000    0.0000000    7.0000000    3.0000000   14.0000000
##        range          sum       median         mean      SE.mean
##   11.0000000 2391.0000000    7.0000000    7.1160714    0.1234165
## CI.mean.0.95          var      std.dev     coef.var
##    0.2427690    5.1178305    2.2622623    0.3179089

stat.desc(props$phonemes_DUTCHPOND)

##      nbr.val     nbr.null       nbr.na          min          max
##  185.0000000    0.0000000  158.0000000    2.0000000   11.0000000
##        range          sum       median         mean      SE.mean
##    9.0000000  996.0000000    5.0000000    5.3837838    0.1433762
```

```
## CI.mean.0.95           var       std.dev      coef.var
##    0.2828727    3.8029965    1.9501273    0.3622224
```

```r
stat.desc(props$phon_neighbours_DUTCHPOND)
```

```
##       nbr.val      nbr.null        nbr.na           min           max
##  336.0000000  106.0000000     7.0000000     0.0000000    42.0000000
##         range           sum        median          mean       SE.mean
##   42.0000000 1536.0000000     1.0000000     4.5714286     0.4126023
## CI.mean.0.95           var       std.dev      coef.var
##    0.8116178   57.2008529     7.5631245     1.6544335
```

```r
stat.desc(props$orth_neighbours_DUTCHPOND)
```

```
##       nbr.val      nbr.null        nbr.na           min           max
##  336.0000000  102.0000000     7.0000000     0.0000000    24.0000000
##         range           sum        median          mean       SE.mean
##   24.0000000 1115.0000000     1.0000000     3.3184524     0.2655425
## CI.mean.0.95           var       std.dev      coef.var
##    0.5223409   23.6923152     4.8674752     1.4667907
```

```r
stat.desc(props$freq_lg10CD_SUBTLEXNL)
```

```
##       nbr.val      nbr.null        nbr.na           min           max
## 290.00000000    0.00000000   53.00000000    0.30000000    3.86000000
##         range           sum        median          mean       SE.mean
##    3.56000000 521.89000000    1.61000000    1.79962069    0.05958115
## CI.mean.0.95           var       std.dev      coef.var
##    0.11726801    1.02947494    1.01463045    0.56380239
```

```r
stat.desc(props$freq_lg10WF_SUBTLEXNL)
```

```
##       nbr.val      nbr.null        nbr.na           min           max
## 290.00000000    0.00000000   53.00000000    0.30000000    4.64000000
##         range           sum        median          mean       SE.mean
##    4.34000000 545.51000000    1.69000000    1.88106897    0.06419049
## CI.mean.0.95           var       std.dev      coef.var
##    0.12634014    1.19492169    1.09312474    0.58111890
```

```r
stat.desc(props$freq_CELEX_lem)
```

```
##       nbr.val      nbr.null        nbr.na           min           max
## 247.00000000   40.00000000   96.00000000    0.00000000    3.08600000
##         range           sum        median          mean       SE.mean
##    3.08600000 266.89000000    0.95400000    1.08052632    0.05081189
## CI.mean.0.95           var       std.dev      coef.var
##    0.10008186    0.63771661    0.79857160    0.73905799
```

```r
stat.desc(props$AoA_Brysbaertetal2014)
```

```
##       nbr.val      nbr.null        nbr.na           min           max
##   233.0000000    0.0000000   110.0000000     3.9100000    14.0800000
```

```
##        range          sum       median         mean      SE.mean
##   10.1700000 1857.6700000    8.0900000    7.9728326    0.1394971
## CI.mean.0.95          var      std.dev     coef.var
##    0.2748431    4.5340523    2.1293314    0.2670734
```

```
stat.desc(props$concrete_Brysbaertetal2014)
```

```
##      nbr.val     nbr.null       nbr.na          min          max
## 233.00000000   0.00000000 110.00000000   1.33000000   4.67000000
##        range          sum       median         mean      SE.mean
##   3.34000000 761.91000000   3.40000000   3.27000000   0.04601172
## CI.mean.0.95          var      std.dev     coef.var
##   0.09065422   0.49327931   0.70233846   0.21478240
```

```
# See and print correlation of all lexical variables:

mat_lexicals_props <- as.matrix(props[c('letters', 'phonemes_DUTCHPOND',
'orth_neighbours_DUTCHPOND', 'phon_neighbours_DUTCHPOND', 'freq_lg10CD_SUBTLEXNL',
'freq_lg10WF_SUBTLEXNL', 'freq_CELEX_lem', 'AoA_Brysbaertetal2014',
'concrete_Brysbaertetal2014')])

rcor.test(mat_lexicals_props, use='complete.obs')
```

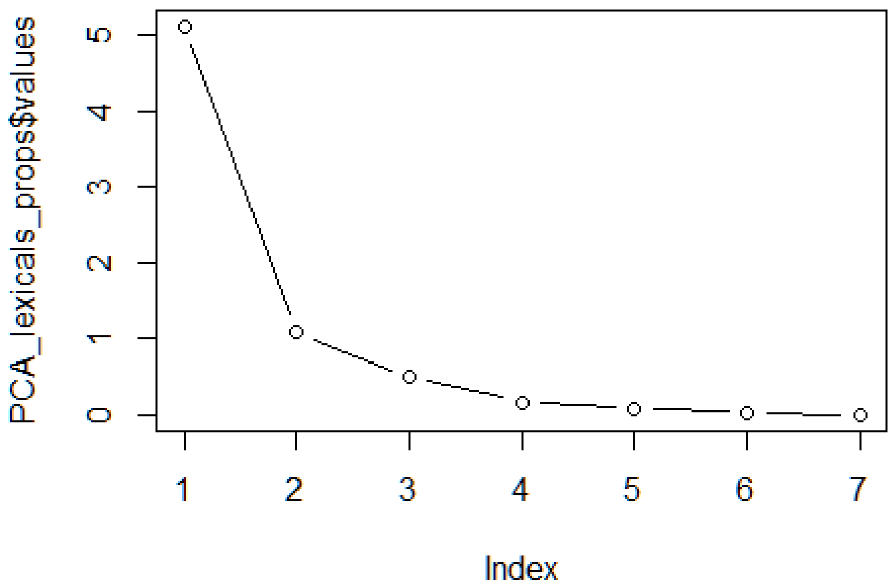
```
##
##                              letters phonemes_DUTCHPOND
## letters                       *****   0.940
## phonemes_DUTCHPOND           <0.001   *****
## orth_neighbours_DUTCHPOND    <0.001  <0.001
## phon_neighbours_DUTCHPOND    <0.001  <0.001
## freq_lg10CD_SUBTLEXNL        <0.001  <0.001
## freq_lg10WF_SUBTLEXNL        <0.001  <0.001
## freq_CELEX_lem               <0.001  <0.001
## AoA_Brysbaertetal2014        <0.001  <0.001
## concrete_Brysbaertetal2014    0.300   0.187
##                              orth_neighbours_DUTCHPOND
## letters                      -0.727
## phonemes_DUTCHPOND           -0.716
## orth_neighbours_DUTCHPOND     *****
## phon_neighbours_DUTCHPOND    <0.001
## freq_lg10CD_SUBTLEXNL        <0.001
## freq_lg10WF_SUBTLEXNL        <0.001
## freq_CELEX_lem               <0.001
## AoA_Brysbaertetal2014        <0.001
## concrete_Brysbaertetal2014    0.168
##                              phon_neighbours_DUTCHPOND freq_lg10CD_SUBTLEXNL
## letters                      -0.703                    -0.508
## phonemes_DUTCHPOND           -0.732                    -0.486
## orth_neighbours_DUTCHPOND     0.895                     0.467
## phon_neighbours_DUTCHPOND     *****                     0.477
## freq_lg10CD_SUBTLEXNL        <0.001                     *****
## freq_lg10WF_SUBTLEXNL        <0.001                    <0.001
```

```
## freq_CELEX_lem              <0.001                    <0.001
## AoA_Brysbaertetal2014      <0.001                    <0.001
## concrete_Brysbaertetal2014  0.060                     0.088
##                            freq_lg10WF_SUBTLEXNL freq_CELEX_lem
## letters                    -0.509                -0.550
## phonemes_DUTCHPOND         -0.486                -0.555
## orth_neighbours_DUTCHPOND   0.470                 0.518
## phon_neighbours_DUTCHPOND   0.478                 0.517
## freq_lg10CD_SUBTLEXNL       0.995                 0.838
## freq_lg10WF_SUBTLEXNL       *****                 0.832
## freq_CELEX_lem             <0.001                 *****
## AoA_Brysbaertetal2014      <0.001                <0.001
## concrete_Brysbaertetal2014  0.097                 0.270
##                            AoA_Brysbaertetal2014
## letters                     0.405
## phonemes_DUTCHPOND          0.457
## orth_neighbours_DUTCHPOND  -0.417
## phon_neighbours_DUTCHPOND  -0.438
## freq_lg10CD_SUBTLEXNL      -0.654
## freq_lg10WF_SUBTLEXNL      -0.646
## freq_CELEX_lem             -0.700
## AoA_Brysbaertetal2014       *****
## concrete_Brysbaertetal2014 <0.001
##                            concrete_Brysbaertetal2014
## letters                    -0.090
## phonemes_DUTCHPOND         -0.090
## orth_neighbours_DUTCHPOND   0.118
## phon_neighbours_DUTCHPOND   0.156
## freq_lg10CD_SUBTLEXNL      -0.166
## freq_lg10WF_SUBTLEXNL      -0.161
## freq_CELEX_lem             -0.100
## AoA_Brysbaertetal2014      -0.254
## concrete_Brysbaertetal2014  *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values

corrs_props = rcor.test(mat_lexicals_props, use='complete.obs')
write.csv(corrs_props$cor.mat, file = "corrs_props.csv",na="") # find table in folder
# (saved just for the manuscript)


# go on to PCA. This does not include age of acquisition or concreteness for a
# better comparison with the English data, and because no correlations > .7 (i.e. half
# of variance explained)

lexicals_props <- props[c('letters', 'phonemes_DUTCHPOND', 'orth_neighbours_DUTCHPOND',
'phon_neighbours_DUTCHPOND', 'freq_lg10CD_SUBTLEXNL', 'freq_lg10WF_SUBTLEXNL',
'freq_CELEX_lem')]
```

```r
str(lexicals_props)

## 'data.frame':    343 obs. of  7 variables:
##  $ letters                   : int  5 11 5 8 7 5 11 10 10 7 ...
##  $ phonemes_DUTCHPOND        : int  4 NA 5 NA 6 4 NA 10 NA 7 ...
##  $ orth_neighbours_DUTCHPOND: int  6 1 1 2 6 1 0 2 1 1 ...
##  $ phon_neighbours_DUTCHPOND: int  4 1 1 2 3 1 0 1 1 1 ...
##  $ freq_lg10CD_SUBTLEXNL    : num  1.53 0.3 1.97 0.3 1.45 1.76 1.26 2.06 0.3 1.79 ...
##  $ freq_lg10WF_SUBTLEXNL    : num  1.71 0.3 2.55 0.3 1.52 1.88 1.26 2.14 0.3 1.84 ...
##  $ freq_CELEX_lem           : num  1.431 NA 0 NA 0.699 ...

# start with PCA for lexical variables, done as in Lynott and Connell (2013)
# Check conditions for a PCA
# Correlations

cor(lexicals_props, use = 'complete.obs')

##                              letters phonemes_DUTCHPOND
## letters                    1.0000000          0.9410143
## phonemes_DUTCHPOND         0.9410143          1.0000000
## orth_neighbours_DUTCHPOND -0.7263499         -0.7167381
## phon_neighbours_DUTCHPOND -0.7002995         -0.7304257
## freq_lg10CD_SUBTLEXNL     -0.5166005         -0.4955464
## freq_lg10WF_SUBTLEXNL     -0.5177211         -0.4961595
## freq_CELEX_lem            -0.5569857         -0.5623170
##                           orth_neighbours_DUTCHPOND
## letters                                  -0.7263499
## phonemes_DUTCHPOND                       -0.7167381
## orth_neighbours_DUTCHPOND                 1.0000000
## phon_neighbours_DUTCHPOND                 0.8960621
## freq_lg10CD_SUBTLEXNL                     0.4715896
## freq_lg10WF_SUBTLEXNL                     0.4739747
## freq_CELEX_lem                            0.5187955
##                           phon_neighbours_DUTCHPOND freq_lg10CD_SUBTLEXNL
## letters                                  -0.7002995            -0.5166005
## phonemes_DUTCHPOND                       -0.7304257            -0.4955464
## orth_neighbours_DUTCHPOND                 0.8960621             0.4715896
## phon_neighbours_DUTCHPOND                 1.0000000             0.4814090
## freq_lg10CD_SUBTLEXNL                     0.4814090             1.0000000
## freq_lg10WF_SUBTLEXNL                     0.4826222             0.9947196
## freq_CELEX_lem                            0.5185262             0.8423256
##                           freq_lg10WF_SUBTLEXNL freq_CELEX_lem
## letters                              -0.5177211     -0.5569857
## phonemes_DUTCHPOND                   -0.4961595     -0.5623170
## orth_neighbours_DUTCHPOND             0.4739747      0.5187955
## phon_neighbours_DUTCHPOND             0.4826222      0.5185262
## freq_lg10CD_SUBTLEXNL                 0.9947196      0.8423256
## freq_lg10WF_SUBTLEXNL                 1.0000000      0.8357985
## freq_CELEX_lem                        0.8357985      1.0000000
```

```r
# Result: all variables fit for PCA, as they have few scores below .3
# The correlations broadly replicate Lynott and Connell.

# now on the raw vars:
cortest.bartlett(lexicals_props)

## R was not square, finding R from data

## $chisq
## [1] 4265.57
##
## $p.value
## [1] 0
##
## $df
## [1] 21

# GOOD: Bartlett's test significant

# KMO: Kaiser-Meyer-Olkin Measure of Sampling Adequacy
lexicals_props_matrix <- cor(lexicals_props, use = 'complete.obs')
KMO(lexicals_props_matrix)

## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = lexicals_props_matrix)
## Overall MSA =  0.78
## MSA for each item =
##                   letters        phonemes_DUTCHPOND
##                      0.76                      0.75
## orth_neighbours_DUTCHPOND phon_neighbours_DUTCHPOND
##                      0.78                      0.77
##     freq_lg10CD_SUBTLEXNL     freq_lg10WF_SUBTLEXNL
##                      0.73                      0.73
##            freq_CELEX_lem
##                      0.97

# Result: .78 = good.

# determinant
det(lexicals_props_matrix)

## [1] 1.766926e-05

# GOOD: above 0.00001

# start off with unrotated PCA

PCA_lexicals_props <- psych::principal(lexicals_props, nfactors = 7, scores = TRUE)
PCA_lexicals_props
```

```
## Principal Components Analysis
## Call: psych::principal(r = lexicals_props, nfactors = 7, scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##                              RC4   RC3   RC1   RC2   RC5   RC6   RC7 h2
## letters                    -0.35  0.86 -0.33 -0.03 -0.04  0.15  0.00  1
## phonemes_DUTCHPOND         -0.23  0.88 -0.38 -0.11  0.04 -0.14  0.00  1
## orth_neighbours_DUTCHPOND  0.28 -0.37  0.86  0.09  0.20 -0.02  0.00  1
## phon_neighbours_DUTCHPOND  0.30 -0.36  0.86  0.06 -0.20  0.02  0.00  1
## freq_lg10CD_SUBTLEXNL      0.94 -0.25  0.24  0.02  0.00  0.00 -0.04  1
## freq_lg10WF_SUBTLEXNL      0.94 -0.25  0.24  0.01  0.00 -0.01  0.04  1
## freq_CELEX_lem             0.78 -0.25  0.29  0.49  0.00  0.00  0.00  1
##                                 u2 com
## letters                    1.1e-16 1.7
## phonemes_DUTCHPOND         1.6e-15 1.6
## orth_neighbours_DUTCHPOND 1.2e-15 1.8
## phon_neighbours_DUTCHPOND 1.4e-15 1.8
## freq_lg10CD_SUBTLEXNL      1.4e-15 1.3
## freq_lg10WF_SUBTLEXNL      1.6e-15 1.3
## freq_CELEX_lem             1.6e-15 2.2
##
##                        RC4  RC3  RC1  RC2  RC5  RC6 RC7
## SS loadings           2.71 1.98 1.92 0.27 0.08 0.04   0
## Proportion Var        0.39 0.28 0.27 0.04 0.01 0.01   0
## Cumulative Var        0.39 0.67 0.94 0.98 0.99 1.00   1
## Proportion Explained  0.39 0.28 0.27 0.04 0.01 0.01   0
## Cumulative Proportion 0.39 0.67 0.94 0.98 0.99 1.00   1
##
## Mean item complexity =  1.7
## Test of the hypothesis that 7 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0
##  with the empirical chi square  0  with prob <  NA
##
## Fit based upon off diagonal values = 1

# By all standards, extract 3 components


# scree analysis
plot(PCA_lexicals_props$values, type = "b")
```




```
# result: again, extract 3 components


PCA_lexicals_props <- psych::principal(lexicals_props, nfactors = 3, rotate =
"varimax", scores = TRUE)

PCA_lexicals_props  # eigenvalues and exp variances good

## Principal Components Analysis
## Call: psych::principal(r = lexicals_props, nfactors = 3, rotate = "varimax",
##     scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##                              RC2   RC1   RC3   h2    u2 com
## letters                    -0.35  0.86 -0.33 0.98 0.024 1.6
## phonemes_DUTCHPOND         -0.25  0.87 -0.39 0.98 0.025 1.6
## orth_neighbours_DUTCHPOND   0.29 -0.37  0.86 0.96 0.042 1.6
## phon_neighbours_DUTCHPOND   0.31 -0.36  0.86 0.96 0.040 1.6
## freq_lg10CD_SUBTLEXNL       0.93 -0.25  0.23 0.98 0.023 1.3
## freq_lg10WF_SUBTLEXNL       0.93 -0.25  0.23 0.98 0.024 1.3
## freq_CELEX_lem              0.85 -0.24  0.31 0.88 0.120 1.4
##
##                        RC2  RC1  RC3
## SS loadings           2.81 1.95 1.94
## Proportion Var        0.40 0.28 0.28
## Cumulative Var        0.40 0.68 0.96
## Proportion Explained  0.42 0.29 0.29
```

```
## Cumulative Proportion 0.42 0.71 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.02
##  with the empirical chi square  6.46  with prob <  0.091
##
## Fit based upon off diagonal values = 1

PCA_lexicals_props$loadings

##
## Loadings:
##                            RC2    RC1    RC3
## letters                    -0.350  0.862 -0.332
## phonemes_DUTCHPOND         -0.246  0.873 -0.392
## orth_neighbours_DUTCHPOND  0.294 -0.368  0.858
## phon_neighbours_DUTCHPOND  0.309 -0.355  0.859
## freq_lg10CD_SUBTLEXNL      0.928 -0.252  0.227
## freq_lg10WF_SUBTLEXNL      0.927 -0.250  0.233
## freq_CELEX_lem             0.852 -0.244  0.309
##
##                  RC2   RC1   RC3
## SS loadings    2.812 1.952 1.940
## Proportion Var 0.402 0.279 0.277
## Cumulative Var 0.402 0.681 0.958

# The PCA replicates Lynott and Connell. Standdized correlation coeffs
# between each PC and its corresponding set of variables are all above .89,
# while the rest of coefficients are all below .33.

PCA_lexicals_props

## Principal Components Analysis
## Call: psych::principal(r = lexicals_props, nfactors = 3, rotate = "varimax",
##     scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##                              RC2   RC1   RC3   h2    u2 com
## letters                    -0.35  0.86 -0.33 0.98 0.024 1.6
## phonemes_DUTCHPOND         -0.25  0.87 -0.39 0.98 0.025 1.6
## orth_neighbours_DUTCHPOND  0.29 -0.37  0.86 0.96 0.042 1.6
## phon_neighbours_DUTCHPOND  0.31 -0.36  0.86 0.96 0.040 1.6
## freq_lg10CD_SUBTLEXNL      0.93 -0.25  0.23 0.98 0.023 1.3
## freq_lg10WF_SUBTLEXNL      0.93 -0.25  0.23 0.98 0.024 1.3
## freq_CELEX_lem             0.85 -0.24  0.31 0.88 0.120 1.4
##
##                        RC2  RC1  RC3
## SS loadings           2.81 1.95 1.94
## Proportion Var        0.40 0.28 0.28
## Cumulative Var        0.40 0.68 0.96
```

```
## Proportion Explained  0.42 0.29 0.29
## Cumulative Proportion 0.42 0.71 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.02
##  with the empirical chi square  6.46  with prob <  0.091
##
## Fit based upon off diagonal values = 1
```

```
# RC1 = length // RC2 = frequency // RC3 = distinctiveness

PCA_lexicals_props$residual
```

```
##                                  letters phonemes_DUTCHPOND
## letters                      0.023517045       -0.023962895
## phonemes_DUTCHPOND          -0.023962895        0.024527905
## orth_neighbours_DUTCHPOND -0.012514682        0.013421737
## phon_neighbours_DUTCHPOND  0.007817494       -0.008192472
## freq_lg10CD_SUBTLEXNL     -0.004433168        0.005936642
## freq_lg10WF_SUBTLEXNL     -0.006253943        0.007759574
## freq_CELEX_lem             0.015885789       -0.019361066
##                           orth_neighbours_DUTCHPOND
## letters                                -0.012514682
## phonemes_DUTCHPOND                      0.013421737
## orth_neighbours_DUTCHPOND               0.041585846
## phon_neighbours_DUTCHPOND              -0.039472453
## freq_lg10CD_SUBTLEXNL                   0.001286716
## freq_lg10WF_SUBTLEXNL                   0.001353789
## freq_CELEX_lem                         -0.004178748
##                           phon_neighbours_DUTCHPOND freq_lg10CD_SUBTLEXNL
## letters                                 0.007817494          -0.004433168
## phonemes_DUTCHPOND                     -0.008192472           0.005936642
## orth_neighbours_DUTCHPOND              -0.039472453           0.001286716
## phon_neighbours_DUTCHPOND               0.040374981           0.006222638
## freq_lg10CD_SUBTLEXNL                   0.006222638           0.023113551
## freq_lg10WF_SUBTLEXNL                   0.006494202           0.020975916
## freq_CELEX_lem                         -0.014031271          -0.050843434
##                           freq_lg10WF_SUBTLEXNL freq_CELEX_lem
## letters                            -0.006253943    0.015885789
## phonemes_DUTCHPOND                  0.007759574   -0.019361066
## orth_neighbours_DUTCHPOND           0.001353789   -0.004178748
## phon_neighbours_DUTCHPOND           0.006494202   -0.014031271
## freq_lg10CD_SUBTLEXNL               0.020975916   -0.050843434
## freq_lg10WF_SUBTLEXNL               0.023956553   -0.052102821
## freq_CELEX_lem                     -0.052102821    0.119626897
```

```
PCA_lexicals_props$fit
```

```
## [1] 0.9985986
```

```r
# Results based on a Kaiser-normalizalized orthogonal (varimax) rotation
# (by default in psych::stats pack). Residuals good: less than half w/ absolute
# values > 0.05. Model fit good, > .90. Communalities (h2) good, all well > .7

props <- cbind(props, PCA_lexicals_props$scores)



# REGRESSION

# standardize (mean-center and scale)
props$s_Auditory <- scale(props$Auditory)
props$s_Haptic <- scale(props$Haptic)
props$s_Visual <- scale(props$Visual)
props$s_freq_lg10CD_SUBTLEXNL <- scale(props$freq_lg10CD_SUBTLEXNL)
props$s_freq_lg10WF_SUBTLEXNL <- scale(props$freq_lg10WF_SUBTLEXNL)
props$s_freq_CELEX_lem <- scale(props$freq_CELEX_lem)
props$s_AoA_Brysbaertetal2014 <- scale(props$AoA_Brysbaertetal2014)
props$s_concrete_Brysbaertetal2014 <- scale(props$concrete_Brysbaertetal2014)
props$s_letters <- scale(props$letters)
props$s_phonemes_DUTCHPOND <- scale(props$phonemes_DUTCHPOND)
props$s_orth_neighbours_DUTCHPOND <- scale(props$orth_neighbours_DUTCHPOND)
props$s_phon_neighbours_DUTCHPOND <- scale(props$phon_neighbours_DUTCHPOND)
props$s_RC1_lexicals <- scale(props$RC1)
props$s_RC2_lexicals <- scale(props$RC2)
props$s_RC3_lexicals <- scale(props$RC3)

# length: letters
fit_letters_props <- lm(props$s_letters ~ props$s_Auditory + props$s_Haptic +
props$s_Visual, data = props)
stat.desc(fit_letters_props$residuals, norm = TRUE)

##                           x
## nbr.val       3.360000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.951464e+00
## max           3.056793e+00
## range         5.008258e+00
## sum           5.710710e-15
## median       -5.660484e-03
## mean          1.694389e-17
## SE.mean       5.381262e-02
## CI.mean.0.95  1.058532e-01
## var           9.729883e-01
## std.dev       9.864017e-01
## coef.var      5.821579e+16
## skewness      2.021250e-01
## skew.2SE      7.596443e-01
```

```r
## kurtosis     -5.827220e-01
## kurt.2SE     -1.098206e+00
## normtest.W    9.848674e-01
## normtest.p    1.348370e-03

# residuals distribution: kurtose. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(props$s_letters)

##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 336    0  1  -0.05   -0.03 1.31 -1.82 3.04  4.86 0.15    -0.63
##      se
## X1 0.05

props$log_s_letters <- log(3 + props$s_letters)

fit_letters_props <- lm(props$log_s_letters ~ props$s_Auditory + props$s_Haptic +
props$s_Visual, data = props)

# check residuals again
stat.desc(fit_letters_props$residuals, norm = TRUE)

##                          x
## nbr.val       3.360000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -9.191181e-01
## max           7.831737e-01
## range         1.702292e+00
## sum           2.737394e-15
## median        5.820411e-02
## mean          8.121029e-18
## SE.mean       1.947699e-02
## CI.mean.0.95  3.831262e-02
## var           1.274627e-01
## std.dev       3.570191e-01
## coef.var      4.396230e+16
## skewness     -4.372009e-01
## skew.2SE     -1.643128e+00
## kurtosis     -5.110083e-01
## kurt.2SE     -9.630531e-01
## normtest.W    9.708083e-01
## normtest.p    2.669646e-06

#  same; go back
fit_letters_props <- lm(props$s_letters ~ props$s_Auditory + props$s_Haptic +
props$s_Visual, data = props)

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
```

```r
# tolerance (pref. > 0.2)
vif(fit_letters_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##         1.390569         1.063636         1.369298

mean(vif(fit_letters_props))

## [1] 1.274501

1/vif(fit_letters_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##        0.7191299        0.9401711        0.7303013

# RESULTS: all good

step_letters_props_AIC <- stepAIC(fit_letters_props, direction="both")

## Start:  AIC=-2.2
## props$s_letters ~ props$s_Auditory + props$s_Haptic + props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC
## - props$s_Visual    1   0.05459 326.01 -4.1460
## <none>                          325.95 -2.2023
## - props$s_Haptic    1   2.28519 328.24 -1.8548
## - props$s_Auditory  1   3.10531 329.06 -1.0164
##
## Step:  AIC=-4.15
## props$s_letters ~ props$s_Auditory + props$s_Haptic
##
##                    Df Sum of Sq    RSS     AIC
## <none>                          326.01 -4.1460
## - props$s_Haptic    1    2.3695 328.38 -3.7127
## + props$s_Visual    1    0.0546 325.95 -2.2023
## - props$s_Auditory  1    4.6444 330.65 -1.3930

step_letters_props_F <- stepAIC(fit_letters_props, direction="both", test="F")

## Start:  AIC=-2.2
## props$s_letters ~ props$s_Auditory + props$s_Haptic + props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC F Value   Pr(F)
## - props$s_Visual    1   0.05459 326.01 -4.1460  0.0556 0.81372
## <none>                          325.95 -2.2023
## - props$s_Haptic    1   2.28519 328.24 -1.8548  2.3276 0.12805
## - props$s_Auditory  1   3.10531 329.06 -1.0164  3.1629 0.07624 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-4.15
## props$s_letters ~ props$s_Auditory + props$s_Haptic
```

```
##
##                     Df Sum of Sq    RSS     AIC F Value  Pr(F)
## <none>                          326.01 -4.1460
## - props$s_Haptic    1    2.3695 328.38 -3.7127  2.4203 0.1207
## + props$s_Visual    1    0.0546 325.95 -2.2023  0.0556 0.8137
## - props$s_Auditory  1    4.6444 330.65 -1.3930  4.7441 0.0301 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_letters_props)
```

```
##
## Call:
## lm(formula = props$s_letters ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual, data = props)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.95146 -0.79178 -0.00566  0.71461  3.05679
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -6.413e-17  5.406e-02   0.000   1.0000
## props$s_Auditory  1.135e-01  6.384e-02   1.778   0.0762 .
## props$s_Haptic   -8.518e-02  5.583e-02  -1.526   0.1280
## props$s_Visual   -1.494e-02  6.335e-02  -0.236   0.8137
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9908 on 332 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.02701,    Adjusted R-squared:  0.01822
## F-statistic: 3.072 on 3 and 332 DF,  p-value: 0.02793
```

```
# length: phonemes_DUTCHPOND
fit_phonemes_DUTCHPOND_props <- lm(props$s_phonemes_DUTCHPOND ~ props$s_Auditory +
props$s_Haptic + props$s_Visual, data = props)
stat.desc(fit_phonemes_DUTCHPOND_props$residuals, norm = TRUE)
```

```
##                          x
## nbr.val       1.850000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.781220e+00
## max           2.893531e+00
## range         4.674751e+00
## sum          -7.140122e-15
## median       -1.586697e-01
## mean         -3.864448e-17
## SE.mean       7.203029e-02
## CI.mean.0.95  1.421115e-01
```

```
## var           9.598471e-01
## std.dev       9.797179e-01
## coef.var     -2.535208e+16
## skewness      3.927364e-01
## skew.2SE      1.099162e+00
## kurtosis     -6.193431e-01
## kurt.2SE     -8.711822e-01
## normtest.W    9.699743e-01
## normtest.p    5.167728e-04
```

```
# residuals distribution: skew. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(props$s_phonemes_DUTCHPOND)
```

```
##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 185    0  1   -0.2   -0.05 1.52 -1.74 2.88  4.62 0.43    -0.54
##      se
## X1 0.07
```

```
props$log_s_phonemes_DUTCHPOND <- log(3 + props$s_phonemes_DUTCHPOND)

fit_phonemes_DUTCHPOND_props <- lm(props$log_s_phonemes_DUTCHPOND ~ props$s_Auditory
 + props$s_Haptic + props$s_Visual, data = props)

# check residuals again
stat.desc(fit_phonemes_DUTCHPOND_props$residuals, norm = TRUE)
```

```
##                         x
## nbr.val       1.850000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -8.201762e-01
## max           7.325980e-01
## range         1.552774e+00
## sum           2.125036e-15
## median        5.242801e-04
## mean          1.148437e-17
## SE.mean       2.459029e-02
## CI.mean.0.95  4.851518e-02
## var           1.118662e-01
## std.dev       3.344641e-01
## coef.var      2.912340e+16
## skewness     -1.073812e-01
## skew.2SE     -3.005307e-01
## kurtosis     -8.521788e-01
## kurt.2SE     -1.198694e+00
## normtest.W    9.803325e-01
## normtest.p    1.039941e-02
```

```
#  worse; back
fit_phonemes_DUTCHPOND_props <- lm(props$s_phonemes_DUTCHPOND ~ props$s_Auditory +
props$s_Haptic + props$s_Visual, data = props)

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_phonemes_DUTCHPOND_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##          1.106936         1.001006         1.107723

mean(vif(fit_phonemes_DUTCHPOND_props))

## [1] 1.071888

1/vif(fit_phonemes_DUTCHPOND_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##         0.9033949        0.9989951        0.9027527

# RESULTS: all good

step_phonemes_DUTCHPOND_props_AIC <- stepAIC(fit_phonemes_DUTCHPOND_props,
direction="both")

## Start:  AIC=-0.58
## props$s_phonemes_DUTCHPOND ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                     Df Sum of Sq    RSS     AIC
## - props$s_Haptic    1    1.3289 177.94 -1.1974
## - props$s_Visual    1    1.4171 178.03 -1.1057
## <none>                          176.61 -0.5842
## - props$s_Auditory  1    5.7163 182.33  3.3087
##
## Step:  AIC=-1.2
## props$s_phonemes_DUTCHPOND ~ props$s_Auditory + props$s_Visual
##
##                     Df Sum of Sq    RSS      AIC
## - props$s_Visual    1    1.5050 179.45 -1.63928
## <none>                          177.94 -1.19741
## + props$s_Haptic    1    1.3289 176.61 -0.58424
## - props$s_Auditory  1    5.8055 183.75  2.74206
##
## Step:  AIC=-1.64
## props$s_phonemes_DUTCHPOND ~ props$s_Auditory
##
##                     Df Sum of Sq    RSS      AIC
## <none>                          179.45 -1.63928
## + props$s_Visual    1    1.5050 177.94 -1.19741
```

```
## + props$s_Haptic    1    1.4168 178.03 -1.10571
## - props$s_Auditory  1    4.5542 184.00  0.99729
```

```r
step_phonemes_DUTCHPOND_props_F <- stepAIC(fit_phonemes_DUTCHPOND_props,
direction="both", test="F")
```

```
## Start:  AIC=-0.58
## props$s_phonemes_DUTCHPOND ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC F Value   Pr(F)
## - props$s_Haptic    1    1.3289 177.94 -1.1974  1.3619 0.24474
## - props$s_Visual    1    1.4171 178.03 -1.1057  1.4524 0.22972
## <none>                          176.61 -0.5842
## - props$s_Auditory  1    5.7163 182.33  3.3087  5.8584 0.01649 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-1.2
## props$s_phonemes_DUTCHPOND ~ props$s_Auditory + props$s_Visual
##
##                    Df Sum of Sq    RSS      AIC F Value   Pr(F)
## - props$s_Visual    1    1.5050 179.45 -1.63928  1.5393 0.21631
## <none>                          177.94 -1.19741
## + props$s_Haptic    1    1.3289 176.61 -0.58424  1.3619 0.24474
## - props$s_Auditory  1    5.8055 183.75  2.74206  5.9380 0.01578 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-1.64
## props$s_phonemes_DUTCHPOND ~ props$s_Auditory
##
##                    Df Sum of Sq    RSS      AIC F Value   Pr(F)
## <none>                          179.45 -1.63928
## + props$s_Visual    1    1.5050 177.94 -1.19741  1.5393 0.21631
## + props$s_Haptic    1    1.4168 178.03 -1.10571  1.4484 0.23035
## - props$s_Auditory  1    4.5542 184.00  0.99729  4.6444 0.03246 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(fit_phonemes_DUTCHPOND_props)
```

```
##
## Call:
## lm(formula = props$s_phonemes_DUTCHPOND ~ props$s_Auditory +
##     props$s_Haptic + props$s_Visual, data = props)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7812 -0.7401 -0.1587  0.8007  2.8935
##
```

```
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.05145    0.08057   0.639   0.5239
## props$s_Auditory    0.25688    0.10613   2.420   0.0165 *
## props$s_Haptic     -0.08477    0.07264  -1.167   0.2447
## props$s_Visual      0.10714    0.08891   1.205   0.2297
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9878 on 181 degrees of freedom
##   (158 observations deleted due to missingness)
## Multiple R-squared:  0.04015,    Adjusted R-squared:  0.02424
## F-statistic: 2.524 on 3 and 181 DF,  p-value: 0.05917
```

```
# distinctiveness: orth neigh size
fit_orth_neighbours_DUTCHPOND_props <- lm(props$s_orth_neighbours_DUTCHPOND ~
props$s_Auditory + props$s_Haptic + props$s_Visual, data = props)
stat.desc(fit_orth_neighbours_DUTCHPOND_props$residuals, norm = TRUE)
```

```
##                          x
## nbr.val       3.360000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -9.012905e-01
## max           4.201627e+00
## range         5.102917e+00
## sum           1.110223e-15
## median       -3.445061e-01
## mean          3.250509e-18
## SE.mean       5.397738e-02
## CI.mean.0.95  1.061773e-01
## var           9.789552e-01
## std.dev       9.894217e-01
## coef.var      3.043898e+17
## skewness      2.005223e+00
## skew.2SE      7.536211e+00
## kurtosis      4.014171e+00
## kurt.2SE      7.565161e+00
## normtest.W    7.527604e-01
## normtest.p    4.267770e-22
```

```
# residuals distribution: skewed and kurtosed. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(props$s_orth_neighbours_DUTCHPOND)
```

```
##    vars   n mean sd median trimmed mad   min  max range skew kurtosis   se
## X1    1 336    0  1  -0.48   -0.22 0.3 -0.68 4.25  4.93 2.06     4.06 0.05
```

```
props$log_s_orth_neighbours_DUTCHPOND <- log(2 + props$s_orth_neighbours_DUTCHPOND)
```

```r
fit_orth_neighbours_DUTCHPOND_props <- lm(props$log_s_orth_neighbours_DUTCHPOND ~
props$s_Auditory + props$s_Haptic + props$s_Visual, data = props)

# check residuals again
stat.desc(fit_orth_neighbours_DUTCHPOND_props$residuals, norm = TRUE)

##                          x
## nbr.val       3.360000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -3.965545e-01
## max           1.203044e+00
## range         1.599599e+00
## sum           2.985459e-15
## median       -1.371519e-01
## mean          8.902735e-18
## SE.mean       2.089711e-02
## CI.mean.0.95  4.110610e-02
## var           1.467276e-01
## std.dev       3.830504e-01
## coef.var      4.302615e+16
## skewness      1.257764e+00
## skew.2SE      4.727041e+00
## kurtosis      7.776239e-01
## kurt.2SE      1.465521e+00
## normtest.W    8.480804e-01
## normtest.p    1.534991e-17

#  quite better

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_orth_neighbours_DUTCHPOND_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##         1.390569         1.063636         1.369298

mean(vif(fit_orth_neighbours_DUTCHPOND_props))

## [1] 1.274501

1/vif(fit_orth_neighbours_DUTCHPOND_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##        0.7191299        0.9401711        0.7303013

# RESULTS: all good

step_orth_neighbours_DUTCHPOND_props_AIC <-
stepAIC(fit_orth_neighbours_DUTCHPOND_props, direction="both")
```

```
## Start:  AIC=-637.85
## props$log_s_orth_neighbours_DUTCHPOND ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC
## - props$s_Visual    1   0.01250 49.166 -639.76
## - props$s_Haptic    1   0.05865 49.212 -639.44
## <none>                          49.154 -637.85
## - props$s_Auditory  1   0.69898 49.853 -635.10
##
## Step:  AIC=-639.76
## props$log_s_orth_neighbours_DUTCHPOND ~ props$s_Auditory + props$s_Haptic
##
##                    Df Sum of Sq    RSS     AIC
## - props$s_Haptic    1   0.05428 49.221 -641.39
## <none>                          49.166 -639.76
## + props$s_Visual    1   0.01250 49.154 -637.85
## - props$s_Auditory  1   0.80432 49.971 -636.31
##
## Step:  AIC=-641.39
## props$log_s_orth_neighbours_DUTCHPOND ~ props$s_Auditory
##
##                    Df Sum of Sq    RSS     AIC
## <none>                          49.221 -641.39
## + props$s_Haptic    1   0.05428 49.166 -639.76
## + props$s_Visual    1   0.00813 49.212 -639.44
## - props$s_Auditory  1   0.95192 50.172 -636.95
```

```
step_orth_neighbours_DUTCHPOND_props_F <-
stepAIC(fit_orth_neighbours_DUTCHPOND_props, direction="both", test="F")
```

```
## Start:  AIC=-637.85
## props$log_s_orth_neighbours_DUTCHPOND ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC F Value  Pr(F)
## - props$s_Visual    1   0.01250 49.166 -639.76  0.0845 0.7715
## - props$s_Haptic    1   0.05865 49.212 -639.44  0.3962 0.5295
## <none>                          49.154 -637.85
## - props$s_Auditory  1   0.69898 49.853 -635.10  4.7212 0.0305 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-639.76
## props$log_s_orth_neighbours_DUTCHPOND ~ props$s_Auditory + props$s_Haptic
##
##                    Df Sum of Sq    RSS     AIC F Value   Pr(F)
## - props$s_Haptic    1   0.05428 49.221 -641.39  0.3676 0.54471
## <none>                          49.166 -639.76
## + props$s_Visual    1   0.01250 49.154 -637.85  0.0845 0.77153
```

```
## - props$s_Auditory  1   0.80432 49.971 -636.31  5.4476 0.02019 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-641.39
## props$log_s_orth_neighbours_DUTCHPOND ~ props$s_Auditory
##
##                     Df Sum of Sq    RSS     AIC F Value   Pr(F)
## <none>                           49.221 -641.39
## + props$s_Haptic    1   0.05428 49.166 -639.76  0.3676 0.54471
## + props$s_Visual    1   0.00813 49.212 -639.44  0.0550 0.81469
## - props$s_Auditory  1   0.95192 50.172 -636.95  6.4595 0.01149 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_orth_neighbours_DUTCHPOND_props)
```

```
##
## Call:
## lm(formula = props$log_s_orth_neighbours_DUTCHPOND ~ props$s_Auditory +
##     props$s_Haptic + props$s_Visual, data = props)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3966 -0.2658 -0.1371  0.1314  1.2030
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.605335   0.020991  28.837   <2e-16 ***
## props$s_Auditory -0.053865   0.024790  -2.173   0.0305 *
## props$s_Haptic    0.013646   0.021681   0.629   0.5295
## props$s_Visual   -0.007149   0.024600  -0.291   0.7715
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3848 on 332 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.0203, Adjusted R-squared:  0.01145
## F-statistic: 2.294 on 3 and 332 DF,  p-value: 0.07784
```

```
# distinctiveness: phon neigh size
fit_phon_neighbours_DUTCHPOND_props <- lm(props$s_phon_neighbours_DUTCHPOND ~
props$s_Auditory + props$s_Haptic + props$s_Visual, data = props)
stat.desc(fit_phon_neighbours_DUTCHPOND_props$residuals, norm = TRUE)
```

```
##                          x
## nbr.val       3.360000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -8.455782e-01
## max           4.812038e+00
```

```
## range         5.657616e+00
## sum           2.758210e-16
## median       -3.634595e-01
## mean          8.576410e-19
## SE.mean       5.382661e-02
## CI.mean.0.95  1.058807e-01
## var           9.734942e-01
## std.dev       9.866581e-01
## coef.var      1.150432e+18
## skewness      2.192935e+00
## skew.2SE      8.241687e+00
## kurtosis      4.900577e+00
## kurt.2SE      9.235695e+00
## normtest.W    7.165177e-01
## normtest.p    1.771880e-23

# residuals distribution: skewed and kurtosed. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(props$s_phon_neighbours_DUTCHPOND)

##    vars   n mean sd median trimmed mad  min  max range skew kurtosis   se
## X1    1 336    0  1  -0.47   -0.24 0.2 -0.6 4.95  5.55 2.27     5.04 0.05

props$log_s_phon_neighbours_DUTCHPOND <- log(2 + props$s_phon_neighbours_DUTCHPOND)

fit_phon_neighbours_DUTCHPOND_props <- lm(props$log_s_phon_neighbours_DUTCHPOND ~
props$s_Auditory + props$s_Haptic + props$s_Visual, data = props)

# check residuals again
stat.desc(fit_phon_neighbours_DUTCHPOND_props$residuals, norm = TRUE)

##                          x
## nbr.val       3.360000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -3.517218e-01
## max           1.277575e+00
## range         1.629297e+00
## sum          -3.259545e-15
## median       -1.356554e-01
## mean         -9.718796e-18
## SE.mean       2.007484e-02
## CI.mean.0.95  3.948862e-02
## var           1.354077e-01
## std.dev       3.679779e-01
## coef.var     -3.786250e+16
## skewness      1.486941e+00
## skew.2SE      5.588356e+00
## kurtosis      1.393719e+00
```

```
## kurt.2SE      2.626621e+00
## normtest.W    8.046259e-01
## normtest.p    8.008540e-20

#  quite better

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_phon_neighbours_DUTCHPOND_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##         1.390569         1.063636         1.369298

mean(vif(fit_phon_neighbours_DUTCHPOND_props))

## [1] 1.274501

1/vif(fit_phon_neighbours_DUTCHPOND_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##        0.7191299        0.9401711        0.7303013

# RESULTS: all good

step_phon_neighbours_DUTCHPOND_props_AIC <-
stepAIC(fit_phon_neighbours_DUTCHPOND_props, direction="both")

## Start:  AIC=-664.82
## props$log_s_phon_neighbours_DUTCHPOND ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC
## - props$s_Visual    1   0.01518 45.377 -666.71
## - props$s_Haptic    1   0.08053 45.442 -666.23
## <none>                          45.362 -664.82
## - props$s_Auditory  1   0.87817 46.240 -660.38
##
## Step:  AIC=-666.71
## props$log_s_phon_neighbours_DUTCHPOND ~ props$s_Auditory + props$s_Haptic
##
##                    Df Sum of Sq    RSS     AIC
## - props$s_Haptic    1   0.07492 45.452 -668.16
## <none>                          45.377 -666.71
## + props$s_Visual    1   0.01518 45.362 -664.82
## - props$s_Auditory  1   1.01291 46.390 -661.29
##
## Step:  AIC=-668.16
## props$log_s_phon_neighbours_DUTCHPOND ~ props$s_Auditory
##
##                    Df Sum of Sq    RSS     AIC
## <none>                          45.452 -668.16
```

```
## + props$s_Haptic    1   0.07492 45.377 -666.71
## + props$s_Visual    1   0.00957 45.442 -666.23
## - props$s_Auditory  1   1.20508 46.657 -661.36
```

```r
step_phon_neighbours_DUTCHPOND_props_F <-
 stepAIC(fit_phon_neighbours_DUTCHPOND_props, direction="both", test="F")
```

```
## Start:  AIC=-664.82
## props$log_s_phon_neighbours_DUTCHPOND ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC F Value  Pr(F)
## - props$s_Visual    1   0.01518 45.377 -666.71  0.1111 0.7391
## - props$s_Haptic    1   0.08053 45.442 -666.23  0.5894 0.4432
## <none>                          45.362 -664.82
## - props$s_Auditory  1   0.87817 46.240 -660.38  6.4273 0.0117 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-666.71
## props$log_s_phon_neighbours_DUTCHPOND ~ props$s_Auditory + props$s_Haptic
##
##                    Df Sum of Sq    RSS     AIC F Value    Pr(F)
## - props$s_Haptic    1   0.07492 45.452 -668.16  0.5498 0.458928
## <none>                          45.377 -666.71
## + props$s_Visual    1   0.01518 45.362 -664.82  0.1111 0.739066
## - props$s_Auditory  1   1.01291 46.390 -661.29  7.4333 0.006742 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-668.16
## props$log_s_phon_neighbours_DUTCHPOND ~ props$s_Auditory
##
##                    Df Sum of Sq    RSS     AIC F Value    Pr(F)
## <none>                          45.452 -668.16
## + props$s_Haptic    1   0.07492 45.377 -666.71  0.5498 0.458928
## + props$s_Visual    1   0.00957 45.442 -666.23  0.0701 0.791309
## - props$s_Auditory  1   1.20508 46.657 -661.36  8.8555 0.003135 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(fit_phon_neighbours_DUTCHPOND_props)
```

```
##
## Call:
## lm(formula = props$log_s_phon_neighbours_DUTCHPOND ~ props$s_Auditory +
##     props$s_Haptic + props$s_Visual, data = props)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3517 -0.2599 -0.1357  0.1060  1.2776
```

```
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.609359   0.020165  30.218   <2e-16 ***
## props$s_Auditory -0.060376   0.023815  -2.535   0.0117 *
## props$s_Haptic    0.015990   0.020828   0.768   0.4432
## props$s_Visual   -0.007878   0.023632  -0.333   0.7391
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3696 on 332 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.02776,    Adjusted R-squared:  0.01897
## F-statistic:  3.16 on 3 and 332 DF,  p-value: 0.02486

# freq: SUBTLEX-NL log-10 CD

fit_freq_lg10CD_SUBTLEXNL_props <- lm(props$s_freq_lg10CD_SUBTLEXNL ~
props$s_Auditory + props$s_Haptic + props$s_Visual, data = props)
stat.desc(fit_freq_lg10CD_SUBTLEXNL_props$residuals, norm = TRUE)

##                          x
## nbr.val       2.900000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.692625e+00
## max           2.415049e+00
## range         4.107675e+00
## sum           1.418483e-14
## median       -2.189748e-01
## mean          4.907464e-17
## SE.mean       5.563757e-02
## CI.mean.0.95  1.095062e-01
## var           8.977063e-01
## std.dev       9.474736e-01
## coef.var      1.930679e+16
## skewness      4.242613e-01
## skew.2SE      1.482369e+00
## kurtosis     -7.702721e-01
## kurt.2SE     -1.350186e+00
## normtest.W    9.623283e-01
## normtest.p    7.531921e-07

# residuals distribution: skew and kurtosed. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(props$s_freq_lg10CD_SUBTLEXNL)

##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 290    0  1  -0.19   -0.04 1.12 -1.48 2.03  3.51 0.35    -1.06
```

```
##      se
## X1 0.06

props$log_s_freq_lg10CD_SUBTLEXNL <- log(3 + props$s_freq_lg10CD_SUBTLEXNL)

fit_freq_lg10CD_SUBTLEXNL_props <- lm(props$log_s_freq_lg10CD_SUBTLEXNL ~
props$s_Auditory + props$s_Haptic + props$s_Visual, data = props)

# check residuals again
stat.desc(fit_freq_lg10CD_SUBTLEXNL_props$residuals, norm = TRUE)

##                          x
## nbr.val       2.900000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -6.999745e-01
## max           7.664373e-01
## range         1.466412e+00
## sum           9.454243e-16
## median       -3.012791e-02
## mean          3.198747e-18
## SE.mean       1.873435e-02
## CI.mean.0.95  3.687306e-02
## var           1.017829e-01
## std.dev       3.190344e-01
## coef.var      9.973731e+16
## skewness      1.133701e-02
## skew.2SE      3.961154e-02
## kurtosis     -8.184529e-01
## kurt.2SE     -1.434641e+00
## normtest.W    9.840065e-01
## normtest.p    2.534289e-03

#  quite better

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_freq_lg10CD_SUBTLEXNL_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##         1.382103         1.039118         1.351219

mean(vif(fit_freq_lg10CD_SUBTLEXNL_props))

## [1] 1.25748

1/vif(fit_freq_lg10CD_SUBTLEXNL_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##        0.7235353        0.9623547        0.7400726
```

```r
# RESULTS: all good

step_freq_lg10CD_SUBTLEXNL_props_AIC <- stepAIC(fit_freq_lg10CD_SUBTLEXNL_props,
direction="both")

## Start:  AIC=-655.63
## props$log_s_freq_lg10CD_SUBTLEXNL ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC
## - props$s_Haptic     1   0.00023 29.416 -657.62
## <none>                           29.415 -655.63
## - props$s_Visual     1   1.01036 30.426 -647.83
## - props$s_Auditory   1   1.03613 30.451 -647.59
##
## Step:  AIC=-657.62
## props$log_s_freq_lg10CD_SUBTLEXNL ~ props$s_Auditory + props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC
## <none>                           29.416 -657.62
## + props$s_Haptic     1   0.00023 29.415 -655.63
## - props$s_Visual     1   1.01031 30.426 -649.83
## - props$s_Auditory   1   1.05585 30.471 -649.40

step_freq_lg10CD_SUBTLEXNL__propsF <- stepAIC(fit_freq_lg10CD_SUBTLEXNL_props,
direction="both", test="F")

## Start:  AIC=-655.63
## props$log_s_freq_lg10CD_SUBTLEXNL ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value    Pr(F)
## - props$s_Haptic     1   0.00023 29.416 -657.62  0.0023 0.962128
## <none>                           29.415 -655.63
## - props$s_Visual     1   1.01036 30.426 -647.83  9.8236 0.001902 **
## - props$s_Auditory   1   1.03613 30.451 -647.59 10.0742 0.001668 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-657.62
## props$log_s_freq_lg10CD_SUBTLEXNL ~ props$s_Auditory + props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value    Pr(F)
## <none>                           29.416 -657.62
## + props$s_Haptic     1   0.00023 29.415 -655.63  0.0023 0.962128
## - props$s_Visual     1   1.01031 30.426 -649.83  9.8574 0.001868 **
## - props$s_Auditory   1   1.05585 30.471 -649.40 10.3017 0.001480 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(fit_freq_lg10CD_SUBTLEXNL_props)
```

```
##
## Call:
## lm(formula = props$log_s_freq_lg10CD_SUBTLEXNL ~ props$s_Auditory +
##     props$s_Haptic + props$s_Visual, data = props)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.69997 -0.25070 -0.03013  0.26364  0.76644
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.0348471  0.0188686  54.845  < 2e-16 ***
## props$s_Auditory -0.0725470  0.0228568  -3.174  0.00167 **
## props$s_Haptic   -0.0009166  0.0192867  -0.048  0.96213
## props$s_Visual    0.0691961  0.0220773   3.134  0.00190 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3207 on 286 degrees of freedom
##   (53 observations deleted due to missingness)
## Multiple R-squared:  0.1252, Adjusted R-squared:  0.116
## F-statistic: 13.64 on 3 and 286 DF,  p-value: 2.415e-08
```

```
# freq: SUBTLEX-NL log-10 WF
fit_freq_lg10WF_SUBTLEXNL_props <- lm(props$s_freq_lg10WF_SUBTLEXNL ~
props$s_Auditory + props$s_Haptic + props$s_Visual, data = props)
stat.desc(fit_freq_lg10WF_SUBTLEXNL_props$residuals, norm = TRUE)
```

```
##                          x
## nbr.val       2.900000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.661721e+00
## max           2.533130e+00
## range         4.194851e+00
## sum           8.913009e-15
## median       -2.272985e-01
## mean          3.085995e-17
## SE.mean       5.576327e-02
## CI.mean.0.95  1.097536e-01
## var           9.017672e-01
## std.dev       9.496143e-01
## coef.var      3.077174e+16
## skewness      5.452976e-01
## skew.2SE      1.905270e+00
## kurtosis     -5.163686e-01
## kurt.2SE     -9.051267e-01
## normtest.W    9.599049e-01
## normtest.p    3.588950e-07
```

```r
# residuals distribution: skew. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(props$s_freq_lg10WF_SUBTLEXNL)

##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 290    0  1  -0.17   -0.06 1.14 -1.45 2.52  3.97 0.46    -0.84
##      se
## X1 0.06

props$log_s_freq_lg10WF_SUBTLEXNL <- log(3 + props$s_freq_lg10WF_SUBTLEXNL)

fit_freq_lg10WF_SUBTLEXNL_props <- lm(props$log_s_freq_lg10WF_SUBTLEXNL ~
props$s_Auditory + props$s_Haptic + props$s_Visual, data = props)

# check residuals again
stat.desc(fit_freq_lg10WF_SUBTLEXNL_props$residuals, norm = TRUE)

##                          x
## nbr.val       2.900000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -6.780013e-01
## max           7.683617e-01
## range         1.446363e+00
## sum          -9.228729e-16
## median       -3.591764e-02
## mean         -3.201621e-18
## SE.mean       1.847247e-02
## CI.mean.0.95  3.635763e-02
## var           9.895732e-02
## std.dev       3.145748e-01
## coef.var     -9.825486e+16
## skewness      9.385398e-02
## skew.2SE      3.279259e-01
## kurtosis     -7.561993e-01
## kurt.2SE     -1.325519e+00
## normtest.W    9.858284e-01
## normtest.p    5.880988e-03

#  quite better

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_freq_lg10WF_SUBTLEXNL_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##         1.382103         1.039118         1.351219

mean(vif(fit_freq_lg10WF_SUBTLEXNL_props))
```

```
## [1] 1.25748

1/vif(fit_freq_lg10WF_SUBTLEXNL_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##        0.7235353        0.9623547        0.7400726

# RESULTS: all good

step_freq_lg10WF_SUBTLEXNL_props_AIC <- stepAIC(fit_freq_lg10WF_SUBTLEXNL_props,
direction="both")

## Start:  AIC=-663.79
## props$log_s_freq_lg10WF_SUBTLEXNL ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC
## - props$s_Haptic     1   0.00042 28.599 -665.79
## <none>                           28.599 -663.79
## - props$s_Visual     1   0.97077 29.569 -656.11
## - props$s_Auditory   1   0.97099 29.570 -656.11
##
## Step:  AIC=-665.79
## props$log_s_freq_lg10WF_SUBTLEXNL ~ props$s_Auditory + props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC
## <none>                           28.599 -665.79
## + props$s_Haptic     1   0.00042 28.599 -663.79
## - props$s_Visual     1   0.97040 29.570 -658.11
## - props$s_Auditory   1   0.98769 29.587 -657.94

step_freq_lg10WF_SUBTLEXNL_props_F <- stepAIC(fit_freq_lg10WF_SUBTLEXNL_props,
direction="both", test="F")

## Start:  AIC=-663.79
## props$log_s_freq_lg10WF_SUBTLEXNL ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value    Pr(F)
## - props$s_Haptic     1   0.00042 28.599 -665.79  0.0042 0.948236
## <none>                           28.599 -663.79
## - props$s_Visual     1   0.97077 29.569 -656.11  9.7082 0.002021 **
## - props$s_Auditory   1   0.97099 29.570 -656.11  9.7103 0.002019 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-665.79
## props$log_s_freq_lg10WF_SUBTLEXNL ~ props$s_Auditory + props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value    Pr(F)
## <none>                           28.599 -665.79
```

```
## + props$s_Haptic    1   0.00042 28.599 -663.79  0.0042 0.948236
## - props$s_Visual    1   0.97040 29.570 -658.11  9.7383 0.001989 **
## - props$s_Auditory  1   0.98769 29.587 -657.94  9.9118 0.001815 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_freq_lg10WF_SUBTLEXNL_props)
```

```
##
## Call:
## lm(formula = props$log_s_freq_lg10WF_SUBTLEXNL ~ props$s_Auditory +
##     props$s_Haptic + props$s_Visual, data = props)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.67800 -0.23858 -0.03592  0.25485  0.76836
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.036229   0.018605  55.697  < 2e-16 ***
## props$s_Auditory -0.070229   0.022537  -3.116  0.00202 **
## props$s_Haptic   -0.001236   0.019017  -0.065  0.94824
## props$s_Visual    0.067827   0.021769   3.116  0.00202 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3162 on 286 degrees of freedom
##   (53 observations deleted due to missingness)
## Multiple R-squared:  0.1224, Adjusted R-squared:  0.1132
## F-statistic:  13.3 on 3 and 286 DF,  p-value: 3.75e-08
```

```
# freq: CELEX log-10 lemma WF
fit_freq_CELEX_lem_props <- lm(props$s_freq_CELEX_lem ~ props$s_Auditory +
props$s_Haptic + props$s_Visual, data = props)
stat.desc(fit_freq_CELEX_lem_props$residuals, norm = TRUE)
```

```
##                         x
## nbr.val       2.470000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.514183e+00
## max           2.434544e+00
## range         3.948728e+00
## sum          -9.034440e-15
## median       -1.388975e-01
## mean         -3.669410e-17
## SE.mean       6.316656e-02
## CI.mean.0.95  1.244163e-01
## var           9.855335e-01
## std.dev       9.927404e-01
## coef.var     -2.705450e+16
```

```
## skewness      3.722828e-01
## skew.2SE      1.201518e+00
## kurtosis     -8.446284e-01
## kurt.2SE     -1.368342e+00
## normtest.W    9.590002e-01
## normtest.p    1.738309e-06
```

```
# residuals distribution: skew and kurtosed. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(props$s_freq_CELEX_lem)
```

```
##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 247    0  1  -0.16   -0.06 1.21 -1.35 2.51  3.86 0.37    -0.84
##      se
## X1 0.06
```

```
props$log_s_freq_CELEX_lem <- log(3 + props$s_freq_CELEX_lem)

fit_freq_CELEX_lem_props <- lm(props$log_s_freq_CELEX_lem ~ props$s_Auditory +
props$s_Haptic + props$s_Visual, data = props)

# check residuals again
stat.desc(fit_freq_CELEX_lem_props$residuals, norm = TRUE)
```

```
##                            x
## nbr.val       2.470000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -6.009444e-01
## max           6.821853e-01
## range         1.283130e+00
## sum          -6.834810e-16
## median        1.096647e-02
## mean         -2.754839e-18
## SE.mean       2.162772e-02
## CI.mean.0.95  4.259913e-02
## var           1.155363e-01
## std.dev       3.399063e-01
## coef.var     -1.233852e+17
## skewness     -1.081089e-01
## skew.2SE     -3.489141e-01
## kurtosis     -1.047055e+00
## kurt.2SE     -1.696284e+00
## normtest.W    9.622853e-01
## normtest.p    4.369197e-06
```

```
#  same; go back
fit_freq_CELEX_lem_props <- lm(props$s_freq_CELEX_lem ~ props$s_Auditory +
props$s_Haptic + props$s_Visual, data = props)
```

```r
# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_freq_CELEX_lem_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##         1.263215         1.013850         1.259341

mean(vif(fit_freq_CELEX_lem_props))

## [1] 1.178802

1/vif(fit_freq_CELEX_lem_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##        0.7916311        0.9863389        0.7940662

# RESULTS: all good

step_freq_CELEX_lem_props_AIC <- stepAIC(fit_freq_CELEX_lem_props, direction="both")

## Start:  AIC=3.4
## props$s_freq_CELEX_lem ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                      Df Sum of Sq    RSS    AIC
## - props$s_Haptic    1    0.2674 242.71 1.6709
## - props$s_Auditory  1    0.3161 242.76 1.7205
## <none>                          242.44 3.3986
## - props$s_Visual    1    3.3187 245.76 4.7568
##
## Step:  AIC=1.67
## props$s_freq_CELEX_lem ~ props$s_Auditory + props$s_Visual
##
##                      Df Sum of Sq    RSS    AIC
## - props$s_Auditory  1    0.3623 243.07 0.0393
## <none>                          242.71 1.6709
## - props$s_Visual    1    3.2356 245.94 2.9420
## + props$s_Haptic    1    0.2674 242.44 3.3986
##
## Step:  AIC=0.04
## props$s_freq_CELEX_lem ~ props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC
## <none>                          243.07 0.03930
## - props$s_Visual    1   2.92911 246.00 0.99797
## + props$s_Auditory  1   0.36229 242.71 1.67087
## + props$s_Haptic    1   0.31352 242.76 1.72050

step_freq_CELEX_lem_props_F <- stepAIC(fit_freq_CELEX_lem_props, direction="both",
test="F")
```

```
## Start:  AIC=3.4
## props$s_freq_CELEX_lem ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                      Df Sum of Sq    RSS    AIC F Value   Pr(F)
## - props$s_Haptic      1    0.2674 242.71 1.6709  0.2680 0.60516
## - props$s_Auditory    1    0.3161 242.76 1.7205  0.3169 0.57402
## <none>                           242.44 3.3986
## - props$s_Visual      1    3.3187 245.76 4.7568  3.3264 0.06941 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=1.67
## props$s_freq_CELEX_lem ~ props$s_Auditory + props$s_Visual
##
##                      Df Sum of Sq    RSS    AIC F Value   Pr(F)
## - props$s_Auditory    1    0.3623 243.07 0.0393  0.3642 0.54673
## <none>                           242.71 1.6709
## - props$s_Visual      1    3.2356 245.94 2.9420  3.2529 0.07253 .
## + props$s_Haptic      1    0.2674 242.44 3.3986  0.2680 0.60516
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=0.04
## props$s_freq_CELEX_lem ~ props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value   Pr(F)
## <none>                           243.07 0.03930
## - props$s_Visual      1   2.92911 246.00 0.99797 2.95236 0.08702 .
## + props$s_Auditory    1   0.36229 242.71 1.67087 0.36422 0.54673
## + props$s_Haptic      1   0.31352 242.76 1.72050 0.31513 0.57507
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(fit_freq_CELEX_lem_props)

##
## Call:
## lm(formula = props$s_freq_CELEX_lem ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual, data = props)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5142 -0.8016 -0.1389  0.7363  2.4345
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.01352    0.06784  -0.199   0.8422
## props$s_Auditory   0.05143    0.09137   0.563   0.5740
## props$s_Haptic    -0.03235    0.06250  -0.518   0.6052
```

```
## props$s_Visual    0.14572    0.07990   1.824   0.0694 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9988 on 243 degrees of freedom
##   (96 observations deleted due to missingness)
## Multiple R-squared:  0.01447,    Adjusted R-squared:  0.002299
## F-statistic: 1.189 on 3 and 243 DF,  p-value: 0.3146

# length: RC1 lexicals
fit_RC1_lexicals_props <- lm(props$s_RC1_lexicals ~ props$s_Auditory + props$s_Haptic
+ props$s_Visual, data = props)
stat.desc(fit_RC1_lexicals_props$residuals, norm = TRUE)

##                            x
## nbr.val       1.700000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.822622e+00
## max           3.222212e+00
## range         5.044834e+00
## sum          -3.788636e-15
## median       -1.386471e-01
## mean         -2.226473e-17
## SE.mean       7.506144e-02
## CI.mean.0.95  1.481788e-01
## var           9.578174e-01
## std.dev       9.786814e-01
## coef.var     -4.395658e+16
## skewness      7.538996e-01
## skew.2SE      2.024025e+00
## kurtosis      2.557611e-01
## kurt.2SE      3.452572e-01
## normtest.W    9.558991e-01
## normtest.p    3.484999e-05

# residuals distribution: skewed. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(props$s_RC1_lexicals)

##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 170    0  1  -0.11   -0.09 0.93 -2.08 3.63  5.71 0.93     0.94
##      se
## X1 0.08

props$log_s_RC1_lexicals_props <- log(4 + props$s_RC1_lexicals)

fit_RC1_lexicals_props <- lm(props$log_s_RC1_lexicals ~ props$s_Auditory +
props$s_Haptic + props$s_Visual, data = props)
```

```
# check residuals again
stat.desc(fit_RC1_lexicals_props$residuals, norm = TRUE)

##                           x
## nbr.val       1.700000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -6.408548e-01
## max           6.186645e-01
## range         1.259519e+00
## sum           6.973588e-16
## median       -9.930210e-03
## mean          4.093142e-18
## SE.mean       1.804027e-02
## CI.mean.0.95  3.561330e-02
## var           5.532672e-02
## std.dev       2.352163e-01
## coef.var      5.746595e+16
## skewness      1.943265e-01
## skew.2SE      5.217161e-01
## kurtosis     -4.515675e-01
## kurt.2SE     -6.095802e-01
## normtest.W    9.897230e-01
## normtest.p    2.566697e-01

# good!

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_RC1_lexicals_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##         1.139419         1.002843         1.141874

mean(vif(fit_RC1_lexicals_props))

## [1] 1.094712

1/vif(fit_RC1_lexicals_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##        0.8776401        0.9971650        0.8757534

# RESULTS: all good

step_RC1_lexicals_props_AIC <- stepAIC(fit_RC1_lexicals_props, direction="both")

## Start:  AIC=-485.07
## props$log_s_RC1_lexicals ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC
```

```
## - props$s_Haptic    1  0.075722 9.4259 -485.70
## - props$s_Visual    1  0.080229 9.4304 -485.62
## <none>                          9.3502 -485.07
## - props$s_Auditory  1  0.246461 9.5967 -482.64
##
## Step:  AIC=-485.7
## props$log_s_RC1_lexicals ~ props$s_Auditory + props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC
## - props$s_Visual    1  0.088911 9.5148 -486.10
## <none>                          9.4259 -485.70
## + props$s_Haptic    1  0.075722 9.3502 -485.07
## - props$s_Auditory  1  0.253554 9.6795 -483.18
##
## Step:  AIC=-486.1
## props$log_s_RC1_lexicals ~ props$s_Auditory
##
##                    Df Sum of Sq    RSS     AIC
## <none>                          9.5148 -486.10
## + props$s_Visual    1  0.088911 9.4259 -485.70
## + props$s_Haptic    1  0.084404 9.4304 -485.62
## - props$s_Auditory  1  0.181715 9.6966 -484.88
```

```
step_RC1_lexicals_props_F <- stepAIC(fit_RC1_lexicals_props, direction="both",
test="F")
```

```
## Start:  AIC=-485.07
## props$log_s_RC1_lexicals ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC F Value   Pr(F)
## - props$s_Haptic    1  0.075722 9.4259 -485.70  1.3443 0.24794
## - props$s_Visual    1  0.080229 9.4304 -485.62  1.4243 0.23439
## <none>                          9.3502 -485.07
## - props$s_Auditory  1  0.246461 9.5967 -482.64  4.3756 0.03798 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-485.7
## props$log_s_RC1_lexicals ~ props$s_Auditory + props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC F Value   Pr(F)
## - props$s_Visual    1  0.088911 9.5148 -486.10  1.5752 0.21120
## <none>                          9.4259 -485.70
## + props$s_Haptic    1  0.075722 9.3502 -485.07  1.3443 0.24794
## - props$s_Auditory  1  0.253554 9.6795 -483.18  4.4922 0.03553 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-486.1
```

```
## props$log_s_RC1_lexicals ~ props$s_Auditory
##
##                      Df Sum of Sq    RSS     AIC F Value   Pr(F)
## <none>                            9.5148 -486.10
## + props$s_Visual    1  0.088911 9.4259 -485.70  1.5752 0.21120
## + props$s_Haptic    1  0.084404 9.4304 -485.62  1.4947 0.22322
## - props$s_Auditory  1  0.181715 9.6966 -484.88  3.2085 0.07506 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_RC1_lexicals_props)
```

```
##
## Call:
## lm(formula = props$log_s_RC1_lexicals ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual, data = props)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.64085 -0.18979 -0.00993  0.15912  0.61866
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.37128    0.02076  66.043   <2e-16 ***
## props$s_Auditory  0.05991    0.02864   2.092    0.038 *
## props$s_Haptic   -0.02067    0.01783  -1.159    0.248
## props$s_Visual    0.02760    0.02313   1.193    0.234
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2373 on 166 degrees of freedom
##   (173 observations deleted due to missingness)
## Multiple R-squared:  0.03572,    Adjusted R-squared:  0.01829
## F-statistic:  2.05 on 3 and 166 DF,  p-value: 0.1089
```

```
# distinctiveness: RC3 lexicals
fit_RC3_lexicals_props <- lm(props$s_RC3_lexicals ~ props$s_Auditory +
props$s_Haptic + props$s_Visual, data = props)
stat.desc(fit_RC3_lexicals_props$residuals, norm = TRUE)
```

```
##                         x
## nbr.val       1.700000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.355748e+00
## max           4.005520e+00
## range         5.361267e+00
## sum           2.810252e-15
## median       -2.582047e-01
## mean          1.657107e-17
## SE.mean       7.489141e-02
```

```
## CI.mean.0.95  1.478432e-01
## var           9.534830e-01
## std.dev       9.764646e-01
## coef.var      5.892585e+16
## skewness      1.582698e+00
## skew.2SE      4.249133e+00
## kurtosis      2.764799e+00
## kurt.2SE      3.732258e+00
## normtest.W    8.564774e-01
## normtest.p    1.279852e-11
```

```
# residuals distribution: skewed and kurtosed. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(props$s_RC3_lexicals)
```

```
##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 170    0  1  -0.37   -0.17 0.57 -1.16 4.03  5.19 1.62     2.42
##      se
## X1 0.08
```

```
props$log_s_RC3_lexicals <- log(3 + props$s_RC3_lexicals)

fit_RC3_lexicals_props <- lm(props$log_s_RC3_lexicals ~ props$s_Auditory +
props$s_Haptic + props$s_Visual, data = props)

# check residuals again
stat.desc(fit_RC3_lexicals_props$residuals, norm = TRUE)
```

```
##                           x
## nbr.val       1.700000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -4.797756e-01
## max           9.240221e-01
## range         1.403798e+00
## sum          -5.325601e-16
## median       -7.295993e-02
## mean         -3.147455e-18
## SE.mean       2.151459e-02
## CI.mean.0.95  4.247196e-02
## var           7.868919e-02
## std.dev       2.805159e-01
## coef.var     -8.912469e+16
## skewness      9.224203e-01
## skew.2SE      2.476459e+00
## kurtosis      4.920814e-01
## kurt.2SE      6.642707e-01
## normtest.W    9.373994e-01
## normtest.p    8.913216e-07
```

```
#  quite better

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_RC3_lexicals_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##         1.139419         1.002843         1.141874

mean(vif(fit_RC3_lexicals_props))

## [1] 1.094712

1/vif(fit_RC3_lexicals_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##        0.8776401        0.9971650        0.8757534

# RESULTS: all good

step_RC3_lexicals_props_AIC <- stepAIC(fit_RC3_lexicals_props, direction="both")

## Start:  AIC=-425.19
## props$log_s_RC3_lexicals ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC
## - props$s_Haptic    1   0.01036 13.309 -427.05
## <none>                          13.299 -425.19
## - props$s_Auditory  1   0.17132 13.470 -425.01
## - props$s_Visual    1   0.59945 13.898 -419.69
##
## Step:  AIC=-427.05
## props$log_s_RC3_lexicals ~ props$s_Auditory + props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC
## <none>                          13.309 -427.05
## - props$s_Auditory  1   0.17356 13.482 -426.85
## + props$s_Haptic    1   0.01036 13.299 -425.19
## - props$s_Visual    1   0.60949 13.918 -421.44

step_RC3_lexicals_props_F <- stepAIC(fit_RC3_lexicals_props, direction="both",
test="F")

## Start:  AIC=-425.19
## props$log_s_RC3_lexicals ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                    Df Sum of Sq    RSS     AIC F Value    Pr(F)
## - props$s_Haptic    1   0.01036 13.309 -427.05  0.1293 0.719589
## <none>                          13.299 -425.19
## - props$s_Auditory  1   0.17132 13.470 -425.01  2.1386 0.145527
```

```
## - props$s_Visual    1   0.59945 13.898 -419.69  7.4828 0.006907 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-427.05
## props$log_s_RC3_lexicals ~ props$s_Auditory + props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value    Pr(F)
## <none>                            13.309 -427.05
## - props$s_Auditory  1   0.17356 13.482 -426.85  2.1779 0.141891
## + props$s_Haptic    1   0.01036 13.299 -425.19  0.1293 0.719589
## - props$s_Visual    1   0.60949 13.918 -421.44  7.6480 0.006323 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_RC3_lexicals_props)
```

```
##
## Call:
## lm(formula = props$log_s_RC3_lexicals ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual, data = props)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.47978 -0.20764 -0.07296  0.13226  0.92402
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.058987   0.024762  42.766  < 2e-16 ***
## props$s_Auditory -0.049950   0.034157  -1.462  0.14553
## props$s_Haptic     0.007645   0.021259   0.360  0.71959
## props$s_Visual    -0.075441   0.027579  -2.735  0.00691 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.283 on 166 degrees of freedom
##   (173 observations deleted due to missingness)
## Multiple R-squared:  0.04616,    Adjusted R-squared:  0.02892
## F-statistic: 2.678 on 3 and 166 DF,  p-value: 0.04882
```

```
# freq: RC2 lexicals
fit_RC2_lexicals_props <- lm(props$s_RC2_lexicals ~ props$s_Auditory + props$s_Haptic
 + props$s_Visual, data = props)
stat.desc(fit_RC2_lexicals_props$residuals, norm = TRUE)
```

```
##                           x
## nbr.val        1.700000e+02
## nbr.null       0.000000e+00
## nbr.na         0.000000e+00
## min           -2.521274e+00
## max            2.428216e+00
```

```
## range          4.949490e+00
## sum           -3.816392e-15
## median        -1.168067e-01
## mean          -2.235880e-17
## SE.mean        7.481168e-02
## CI.mean.0.95   1.476858e-01
## var            9.514538e-01
## std.dev        9.754249e-01
## coef.var      -4.362600e+16
## skewness       2.099120e-01
## skew.2SE       5.635592e-01
## kurtosis      -7.566392e-01
## kurt.2SE      -1.021403e+00
## normtest.W     9.788257e-01
## normtest.p     1.063544e-02

# residuals distribution: kurtosed. Raw scores/2.SE < 1
# have to log-transform DV and re-run regression

psych::describe(props$s_RC2_lexicals)

##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 170    0  1  -0.02   -0.03 1.15 -2.37 2.29  4.66  0.2    -0.67
##      se
## X1 0.08

props$log_s_RC2_lexicals <- log(3 + props$s_RC2_lexicals)

fit_RC2_lexicals_props <- lm(props$log_s_RC2_lexicals ~ props$s_Auditory +
props$s_Haptic + props$s_Visual, data = props)

# check residuals again
stat.desc(fit_RC2_lexicals_props$residuals, norm = TRUE)

##                            x
## nbr.val        1.700000e+02
## nbr.null       0.000000e+00
## nbr.na         0.000000e+00
## min           -1.548065e+00
## max            7.294142e-01
## range          2.277479e+00
## sum            8.968520e-16
## median         1.992494e-02
## mean           5.315142e-18
## SE.mean        2.746269e-02
## CI.mean.0.95   5.421412e-02
## var            1.282139e-01
## std.dev        3.580697e-01
## coef.var       6.736786e+16
## skewness      -6.661503e-01
```

```
## skew.2SE     -1.788441e+00
## kurtosis      9.513419e-01
## kurt.2SE      1.284236e+00
## normtest.W    9.666991e-01
## normtest.p    4.244338e-04

# worse; back
fit_RC2_lexicals_props <- lm(props$s_RC2_lexicals ~ props$s_Auditory + props$s_Haptic
 + props$s_Visual, data = props)

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_RC2_lexicals_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##          1.139419         1.002843         1.141874

mean(vif(fit_RC2_lexicals_props))

## [1] 1.094712

1/vif(fit_RC2_lexicals_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##         0.8776401        0.9971650        0.8757534

# RESULTS: all good

step_RC2_lexicals_props_AIC <- stepAIC(fit_RC2_lexicals_props, direction="both")

## Start:  AIC=-1.46
## props$s_RC2_lexicals ~ props$s_Auditory + props$s_Haptic + props$s_Visual
##
##                     Df Sum of Sq    RSS     AIC
## - props$s_Haptic     1    0.8921 161.69 -2.5223
## <none>                           160.80 -1.4629
## - props$s_Visual     1    3.7779 164.57  0.4851
## - props$s_Auditory   1    5.6714 166.47  2.4298
##
## Step:  AIC=-2.52
## props$s_RC2_lexicals ~ props$s_Auditory + props$s_Visual
##
##                     Df Sum of Sq    RSS      AIC
## <none>                           161.69 -2.52228
## + props$s_Haptic     1    0.8921 160.80 -1.46286
## - props$s_Visual     1    3.9852 165.67 -0.38301
## - props$s_Auditory   1    5.7890 167.48  1.45786

step_RC2_lexicals_props_F <- stepAIC(fit_RC2_lexicals_props, direction="both",
test="F")
```

```
## Start:  AIC=-1.46
## props$s_RC2_lexicals ~ props$s_Auditory + props$s_Haptic + props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value   Pr(F)
## - props$s_Haptic    1    0.8921 161.69 -2.5223  0.9210 0.33861
## <none>                           160.80 -1.4629
## - props$s_Visual    1    3.7779 164.57  0.4851  3.9002 0.04994 *
## - props$s_Auditory  1    5.6714 166.47  2.4298  5.8549 0.01661 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-2.52
## props$s_RC2_lexicals ~ props$s_Auditory + props$s_Visual
##
##                      Df Sum of Sq    RSS      AIC F Value   Pr(F)
## <none>                           161.69 -2.52228
## + props$s_Haptic    1    0.8921 160.80 -1.46286  0.9210 0.33861
## - props$s_Visual    1    3.9852 165.67 -0.38301  4.1161 0.04406 *
## - props$s_Auditory  1    5.7890 167.48  1.45786  5.9792 0.01551 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(fit_RC2_lexicals_props)

##
## Call:
## lm(formula = props$s_RC2_lexicals ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual, data = props)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5213 -0.8128 -0.1168  0.7561  2.4282
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.04703    0.08610   0.546   0.5856
## props$s_Auditory  0.28739    0.11877   2.420   0.0166 *
## props$s_Haptic   -0.07094    0.07392  -0.960   0.3386
## props$s_Visual    0.18939    0.09590   1.975   0.0499 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9842 on 166 degrees of freedom
##   (173 observations deleted due to missingness)
## Multiple R-squared:  0.04855,    Adjusted R-squared:  0.03135
## F-statistic: 2.823 on 3 and 166 DF,  p-value: 0.04046

# additional var: age of acquisition
fit_AoA_Brysbaertetal2014_props <- lm(props$s_AoA_Brysbaertetal2014 ~
```

```r
props$s_Auditory + props$s_Haptic + props$s_Visual, data = props)
stat.desc(fit_AoA_Brysbaertetal2014_props$residuals, norm = TRUE)

##                          x
## nbr.val       2.330000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -2.112829e+00
## max           2.593592e+00
## range         4.706422e+00
## sum           2.532696e-16
## median        3.452361e-02
## mean          1.144694e-18
## SE.mean       6.406750e-02
## CI.mean.0.95  1.262285e-01
## var           9.563821e-01
## std.dev       9.779479e-01
## coef.var      8.543312e+17
## skewness      7.860513e-02
## skew.2SE      2.464866e-01
## kurtosis     -4.681725e-01
## kurt.2SE     -7.370869e-01
## normtest.W    9.881971e-01
## normtest.p    5.248873e-02

# residuals distribution: good

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_AoA_Brysbaertetal2014_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##         1.292539         1.013180         1.284802

mean(vif(fit_AoA_Brysbaertetal2014_props))

## [1] 1.19684

1/vif(fit_AoA_Brysbaertetal2014_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##        0.7736713        0.9869918        0.7783300

# RESULTS: all good

step_AoA_Brysbaertetal2014_props_AIC <- stepAIC(fit_AoA_Brysbaertetal2014_props,
direction="both")

## Start:  AIC=-3.39
## props$s_AoA_Brysbaertetal2014 ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
```

```
##                       Df Sum of Sq    RSS     AIC
## - props$s_Haptic     1    0.0101 221.89 -5.3829
## - props$s_Auditory   1    0.1463 222.03 -5.2398
## <none>                           221.88 -3.3934
## - props$s_Visual     1    6.7408 228.62  1.5798
##
## Step:  AIC=-5.38
## props$s_AoA_Brysbaertetal2014 ~ props$s_Auditory + props$s_Visual
##
##                       Df Sum of Sq    RSS     AIC
## - props$s_Auditory   1    0.1538 222.04 -7.2214
## <none>                           221.89 -5.3829
## + props$s_Haptic     1    0.0101 221.88 -3.3934
## - props$s_Visual     1    6.7630 228.65 -0.3874
##
## Step:  AIC=-7.22
## props$s_AoA_Brysbaertetal2014 ~ props$s_Visual
##
##                       Df Sum of Sq    RSS     AIC
## <none>                           222.04 -7.2214
## + props$s_Auditory   1    0.1538 221.89 -5.3829
## + props$s_Haptic     1    0.0176 222.03 -5.2398
## - props$s_Visual     1    9.9555 232.00  0.9978
```

```
step_AoA_Brysbaertetal2014_props_F <- stepAIC(fit_AoA_Brysbaertetal2014_props,
direction="both", test="F")
```

```
## Start:  AIC=-3.39
## props$s_AoA_Brysbaertetal2014 ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                       Df Sum of Sq    RSS     AIC F Value    Pr(F)
## - props$s_Haptic     1    0.0101 221.89 -5.3829  0.0104 0.918933
## - props$s_Auditory   1    0.1463 222.03 -5.2398  0.1510 0.697929
## <none>                           221.88 -3.3934
## - props$s_Visual     1    6.7408 228.62  1.5798  6.9571 0.008921 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-5.38
## props$s_AoA_Brysbaertetal2014 ~ props$s_Auditory + props$s_Visual
##
##                       Df Sum of Sq    RSS     AIC F Value    Pr(F)
## - props$s_Auditory   1    0.1538 222.04 -7.2214  0.1594 0.690035
## <none>                           221.89 -5.3829
## + props$s_Haptic     1    0.0101 221.88 -3.3934  0.0104 0.918933
## - props$s_Visual     1    6.7630 228.65 -0.3874  7.0101 0.008666 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Step:  AIC=-7.22
## props$s_AoA_Brysbaertetal2014 ~ props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value    Pr(F)
## <none>                            222.04 -7.2214
## + props$s_Auditory  1    0.1538 221.89 -5.3829  0.1594 0.690035
## + props$s_Haptic    1    0.0176 222.03 -5.2398  0.0182 0.892810
## - props$s_Visual    1    9.9555 232.00  0.9978 10.3570 0.001475 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(fit_AoA_Brysbaertetal2014_props)
```

```
##
## Call:
## lm(formula = props$s_AoA_Brysbaertetal2014 ~ props$s_Auditory +
##     props$s_Haptic + props$s_Visual, data = props)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.11283 -0.73187  0.03452  0.64255  2.59359
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.054540   0.067911   0.803  0.42274
## props$s_Auditory    0.035560   0.091508   0.389  0.69793
## props$s_Haptic     -0.006453   0.063334  -0.102  0.91893
## props$s_Visual     -0.209357   0.079373  -2.638  0.00892 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9843 on 229 degrees of freedom
##   (110 observations deleted due to missingness)
## Multiple R-squared:  0.04362,    Adjusted R-squared:  0.03109
## F-statistic: 3.481 on 3 and 229 DF,  p-value: 0.01667
```

```r
# additional var: concreteness
fit_concrete_Brysbaertetal2014_props <- lm(props$s_concrete_Brysbaertetal2014 ~
props$s_Auditory + props$s_Haptic + props$s_Visual, data = props)
stat.desc(fit_concrete_Brysbaertetal2014_props$residuals, norm = TRUE)
```

```
##                           x
## nbr.val       2.330000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -2.594543e+00
## max           1.852867e+00
## range         4.447410e+00
## sum          -2.477185e-14
## median        1.768909e-01
## mean         -1.062963e-16
```

```
## SE.mean       6.295441e-02
## CI.mean.0.95  1.240354e-01
## var           9.234389e-01
## std.dev       9.609573e-01
## coef.var     -9.040369e+15
## skewness     -4.207250e-01
## skew.2SE     -1.319291e+00
## kurtosis     -3.351918e-01
## kurt.2SE     -5.277232e-01
## normtest.W    9.790502e-01
## normtest.p    1.581048e-03

# residuals distribution: skew. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(props$s_concrete_Brysbaertetal2014)

##    vars   n mean sd median trimmed  mad   min  max range  skew kurtosis
## X1    1 233    0  1   0.19    0.04 0.99 -2.76 1.99  4.76 -0.37    -0.41
##      se
## X1 0.07

props$log_s_concrete_Brysbaertetal2014 <- log(4 + props$s_concrete_Brysbaertetal2014)

fit_concrete_Brysbaertetal2014_props <- lm(props$log_s_concrete_Brysbaertetal2014 ~
props$s_Auditory + props$s_Haptic + props$s_Visual, data = props)

# check residuals again
stat.desc(fit_concrete_Brysbaertetal2014_props$residuals, norm = TRUE)

##                          x
## nbr.val       2.330000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.084648e+00
## max           4.612322e-01
## range         1.545880e+00
## sum          -4.888451e-15
## median        8.076837e-02
## mean         -2.101351e-17
## SE.mean       1.814043e-02
## CI.mean.0.95  3.574103e-02
## var           7.667449e-02
## std.dev       2.769016e-01
## coef.var     -1.317731e+16
## skewness     -1.152705e+00
## skew.2SE     -3.614602e+00
## kurtosis      1.563516e+00
## kurt.2SE      2.461586e+00
```

```
## normtest.W    9.220715e-01
## normtest.p    1.003247e-09

# worse; back
fit_concrete_Brysbaertetal2014_props <- lm(props$s_concrete_Brysbaertetal2014 ~
props$s_Auditory + props$s_Haptic + props$s_Visual, data = props)

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_concrete_Brysbaertetal2014_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##         1.292539         1.013180         1.284802

mean(vif(fit_concrete_Brysbaertetal2014_props))

## [1] 1.19684

1/vif(fit_concrete_Brysbaertetal2014_props)

## props$s_Auditory   props$s_Haptic   props$s_Visual
##        0.7736713        0.9869918        0.7783300

# RESULTS: all good

step_concrete_Brysbaertetal2014_props_AIC <-
stepAIC(fit_concrete_Brysbaertetal2014_props, direction="both")

## Start:  AIC=-11.56
## props$s_concrete_Brysbaertetal2014 ~ props$s_Auditory + props$s_Haptic +
##     props$s_Visual
##
##                      Df Sum of Sq    RSS      AIC
## - props$s_Haptic    1    1.2182 215.46 -12.2396
## <none>                            214.24 -11.5607
## - props$s_Auditory  1    3.2391 217.48 -10.0643
## - props$s_Visual    1    5.0553 219.29  -8.1266
##
## Step:  AIC=-12.24
## props$s_concrete_Brysbaertetal2014 ~ props$s_Auditory + props$s_Visual
##
##                      Df Sum of Sq    RSS     AIC
## <none>                            215.46 -12.240
## + props$s_Haptic    1    1.2182 214.24 -11.561
## - props$s_Auditory  1    3.6026 219.06 -10.376
## - props$s_Visual    1    5.2123 220.67  -8.670

step_concrete_Brysbaertetal2014_props_F <-
stepAIC(fit_concrete_Brysbaertetal2014_props, direction="both", test="F")

## Start:  AIC=-11.56
## props$s_concrete_Brysbaertetal2014 ~ props$s_Auditory + props$s_Haptic +
```

```
##     props$s_Visual
##
##                     Df Sum of Sq    RSS      AIC F Value   Pr(F)
## - props$s_Haptic    1    1.2182 215.46 -12.2396  1.3022 0.25501
## <none>                          214.24 -11.5607
## - props$s_Auditory  1    3.2391 217.48 -10.0643  3.4623 0.06406 .
## - props$s_Visual    1    5.0553 219.29  -8.1266  5.4036 0.02097 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-12.24
## props$s_concrete_Brysbaertetal2014 ~ props$s_Auditory + props$s_Visual
##
##                     Df Sum of Sq    RSS     AIC F Value   Pr(F)
## <none>                          215.46 -12.240
## + props$s_Haptic    1    1.2182 214.24 -11.561  1.3022 0.25501
## - props$s_Auditory  1    3.6026 219.06 -10.376  3.8457 0.05108 .
## - props$s_Visual    1    5.2123 220.67  -8.670  5.5641 0.01917 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_concrete_Brysbaertetal2014_props)
```

```
##
## Call:
## lm(formula = props$s_concrete_Brysbaertetal2014 ~ props$s_Auditory +
##     props$s_Haptic + props$s_Visual, data = props)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5945 -0.6710  0.1769  0.6787  1.8529
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.08550    0.06673  -1.281   0.2014
## props$s_Auditory -0.16731    0.08992  -1.861   0.0641 .
## props$s_Haptic    0.07102    0.06223   1.141   0.2550
## props$s_Visual    0.18130    0.07799   2.325   0.0210 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9672 on 229 degrees of freedom
##   (110 observations deleted due to missingness)
## Multiple R-squared:  0.07656,    Adjusted R-squared:  0.06446
## F-statistic: 6.329 on 3 and 229 DF,  p-value: 0.0003852
```

```
# RESULTS: iconicity properties:
# Auditory strength either was the strongest predictor or presented an opposite
# polarity from the main predictor. This held for all lexical DVs except age of
# acquisition.
```

```r
# ___________________________________________________________________________

# Iconicity within concepts alone, as in Lynott and Connell (2013)

concs <- all[all$cat == 'conc' & c(all$normed == 'Dutch' | all$normed == 'Dut_Eng'),]
nrow(concs)

## [1] 411

# There aren't lexical data for every single word.
# Percentage of concepts per lexical variable (from items w/ Dutch norms)
describe(complete.cases(concs[complete.cases(concs$Exclusivity),]
$phonemes_DUTCHPOND))

## complete.cases(concs[complete.cases(concs$Exclusivity), ]$phonemes_DUTCHPOND)
##        n missing  unique
##      411       0       2
##
## FALSE (22, 5%), TRUE (389, 95%)

describe(complete.cases(concs[complete.cases(concs$Exclusivity),]
$phon_neighbours_DUTCHPOND))

## complete.cases(concs[complete.cases(concs$Exclusivity), ]$phon_neighbours_DUTCHPOND)
##        n missing  unique   value
##      411       0       1    TRUE

describe(complete.cases(concs[complete.cases(concs$Exclusivity),]
$orth_neighbours_DUTCHPOND))

## complete.cases(concs[complete.cases(concs$Exclusivity), ]$orth_neighbours_DUTCHPOND)
##        n missing  unique   value
##      411       0       1    TRUE

describe(complete.cases(concs[complete.cases(concs$Exclusivity),]
$freq_lg10CD_SUBTLEXNL))

## complete.cases(concs[complete.cases(concs$Exclusivity), ]$freq_lg10CD_SUBTLEXNL)
##        n missing  unique
##      411       0       2
##
## FALSE (4, 1%), TRUE (407, 99%)

describe(complete.cases(concs[complete.cases(concs$Exclusivity),]
$freq_lg10WF_SUBTLEXNL))

## complete.cases(concs[complete.cases(concs$Exclusivity), ]$freq_lg10WF_SUBTLEXNL)
##        n missing  unique
##      411       0       2
```

```
##
## FALSE (4, 1%), TRUE (407, 99%)
```

```
describe(complete.cases(concs[complete.cases(concs$Exclusivity),]
$freq_CELEX_lem))
```

```
## complete.cases(concs[complete.cases(concs$Exclusivity), ]$freq_CELEX_lem)
##        n missing  unique
##      411       0       2
##
## FALSE (12, 3%), TRUE (399, 97%)
```

```
describe(complete.cases(concs[complete.cases(concs$Exclusivity),]
$AoA_Brysbaertetal2014))
```

```
## complete.cases(concs[complete.cases(concs$Exclusivity), ]$AoA_Brysbaertetal2014)
##        n missing  unique
##      411       0       2
##
## FALSE (13, 3%), TRUE (398, 97%)
```

```
describe(complete.cases(concs[complete.cases(concs$Exclusivity),]
$concrete_Brysbaertetal2014))
```

```
## complete.cases(concs[complete.cases(concs$Exclusivity), ]$concrete_Brysbaertetal2014)
##        n missing  unique
##      411       0       2
##
## FALSE (13, 3%), TRUE (398, 97%)
```

```
# M, SD
stat.desc(concs$letters)
```

```
##      nbr.val     nbr.null       nbr.na          min          max
##  411.0000000    0.0000000    0.0000000    3.0000000   17.0000000
##        range          sum       median         mean      SE.mean
##   14.0000000 2759.0000000    6.0000000    6.7128954    0.1254065
## CI.mean.0.95          var      std.dev     coef.var
##    0.2465199    6.4637114    2.5423830    0.3787312
```

```
stat.desc(concs$phonemes_DUTCHPOND)
```

```
##      nbr.val     nbr.null       nbr.na          min          max
##  389.0000000    0.0000000   22.0000000    2.0000000   15.0000000
##        range          sum       median         mean      SE.mean
##   13.0000000 2265.0000000    6.0000000    5.8226221    0.1104353
## CI.mean.0.95          var      std.dev     coef.var
##    0.2171266    4.7442292    2.1781252    0.3740798
```

```
stat.desc(concs$phon_neighbours_DUTCHPOND)
```

```
##       nbr.val     nbr.null       nbr.na          min          max
##  411.0000000  139.0000000    0.0000000    0.0000000   49.0000000
##        range          sum       median         mean      SE.mean
##   49.0000000 2273.0000000    1.0000000    5.5304136    0.4080256
## CI.mean.0.95          var      std.dev     coef.var
##    0.8020832   68.4252923    8.2719582    1.4957214
```

```r
stat.desc(concs$orth_neighbours_DUTCHPOND)
```

```
##       nbr.val     nbr.null       nbr.na          min          max
##  411.0000000  143.0000000    0.0000000    0.0000000   32.0000000
##        range          sum       median         mean      SE.mean
##   32.0000000 1658.0000000    1.0000000    4.0340633    0.2825354
## CI.mean.0.95          var      std.dev     coef.var
##    0.5553987   32.8085930    5.7278786    1.4198782
```

```r
stat.desc(concs$freq_lg10CD_SUBTLEXNL)
```

```
##       nbr.val     nbr.null       nbr.na          min          max
## 4.070000e+02 0.000000e+00 4.000000e+00 6.000000e-01 3.900000e+00
##        range          sum       median         mean      SE.mean
## 3.300000e+00 1.080720e+03 2.700000e+00 2.655332e+00 3.275722e-02
## CI.mean.0.95          var      std.dev     coef.var
## 6.439494e-02 4.367254e-01 6.608521e-01 2.488774e-01
```

```r
stat.desc(concs$freq_lg10WF_SUBTLEXNL)
```

```
##       nbr.val     nbr.null       nbr.na          min          max
## 4.070000e+02 0.000000e+00 4.000000e+00 6.000000e-01 4.770000e+00
##        range          sum       median         mean      SE.mean
## 4.170000e+00 1.157980e+03 2.850000e+00 2.845160e+00 3.775280e-02
## CI.mean.0.95          var      std.dev     coef.var
## 7.421538e-02 5.800866e-01 7.616342e-01 2.676947e-01
```

```r
stat.desc(concs$freq_CELEX_lem)
```

```
##       nbr.val     nbr.null       nbr.na          min          max
## 399.00000000  11.00000000  12.00000000   0.00000000   3.18800000
##        range          sum       median         mean      SE.mean
##   3.18800000 612.79100000   1.55600000   1.53581704   0.03185900
## CI.mean.0.95          var      std.dev     coef.var
##   0.06263295   0.40498325   0.63638294   0.41436117
```

```r
stat.desc(concs$AoA_Brysbaertetal2014)
```

```
##       nbr.val     nbr.null       nbr.na          min          max
##  398.0000000    0.0000000   13.0000000    3.4100000   13.1800000
##        range          sum       median         mean      SE.mean
##    9.7700000 3210.3200000    8.1500000    8.0661307    0.1086208
## CI.mean.0.95          var      std.dev     coef.var
##    0.2135439    4.6957941    2.1669781    0.2686515
```

```r
stat.desc(concs$concrete_Brysbaertetal2014)

##      nbr.val     nbr.null       nbr.na          min          max
## 3.980000e+02 0.000000e+00 1.300000e+01 1.200000e+00 5.000000e+00
##        range          sum       median         mean      SE.mean
## 3.800000e+00 1.203330e+03 2.800000e+00 3.023442e+00 5.356266e-02
## CI.mean.0.95          var      std.dev     coef.var
## 1.053019e-01 1.141845e+00 1.068572e+00 3.534288e-01

# See and print correlation of all lexical variables:

mat_lexicals_concs <- as.matrix(concs[c('letters', 'phonemes_DUTCHPOND',
'orth_neighbours_DUTCHPOND', 'phon_neighbours_DUTCHPOND', 'freq_lg10CD_SUBTLEXNL',
'freq_lg10WF_SUBTLEXNL', 'freq_CELEX_lem', 'AoA_Brysbaertetal2014',
'concrete_Brysbaertetal2014')])

rcor.test(mat_lexicals_concs, use='complete.obs')

##
##                              letters phonemes_DUTCHPOND
## letters                       *****   0.942
## phonemes_DUTCHPOND           <0.001   *****
## orth_neighbours_DUTCHPOND    <0.001  <0.001
## phon_neighbours_DUTCHPOND    <0.001  <0.001
## freq_lg10CD_SUBTLEXNL        <0.001  <0.001
## freq_lg10WF_SUBTLEXNL        <0.001  <0.001
## freq_CELEX_lem               <0.001  <0.001
## AoA_Brysbaertetal2014        <0.001  <0.001
## concrete_Brysbaertetal2014 <0.001  <0.001
##                              orth_neighbours_DUTCHPOND
## letters                      -0.647
## phonemes_DUTCHPOND           -0.617
## orth_neighbours_DUTCHPOND     *****
## phon_neighbours_DUTCHPOND    <0.001
## freq_lg10CD_SUBTLEXNL        <0.001
## freq_lg10WF_SUBTLEXNL        <0.001
## freq_CELEX_lem               <0.001
## AoA_Brysbaertetal2014        <0.001
## concrete_Brysbaertetal2014 <0.001
##                              phon_neighbours_DUTCHPOND freq_lg10CD_SUBTLEXNL
## letters                      -0.630                    -0.364
## phonemes_DUTCHPOND           -0.633                    -0.362
## orth_neighbours_DUTCHPOND     0.879                     0.329
## phon_neighbours_DUTCHPOND     *****                     0.349
## freq_lg10CD_SUBTLEXNL        <0.001                     *****
## freq_lg10WF_SUBTLEXNL        <0.001                    <0.001
## freq_CELEX_lem               <0.001                    <0.001
## AoA_Brysbaertetal2014        <0.001                    <0.001
## concrete_Brysbaertetal2014 <0.001                     0.378
##                              freq_lg10WF_SUBTLEXNL freq_CELEX_lem
```

```
## letters                      -0.381                -0.212
## phonemes_DUTCHPOND           -0.369                -0.237
## orth_neighbours_DUTCHPOND     0.338                 0.201
## phon_neighbours_DUTCHPOND     0.352                 0.220
## freq_lg10CD_SUBTLEXNL         0.987                 0.776
## freq_lg10WF_SUBTLEXNL         *****                 0.757
## freq_CELEX_lem               <0.001                 *****
## AoA_Brysbaertetal2014        <0.001                <0.001
## concrete_Brysbaertetal2014    0.157                 0.028
##                              AoA_Brysbaertetal2014
## letters                       0.491
## phonemes_DUTCHPOND            0.513
## orth_neighbours_DUTCHPOND    -0.467
## phon_neighbours_DUTCHPOND    -0.437
## freq_lg10CD_SUBTLEXNL        -0.585
## freq_lg10WF_SUBTLEXNL        -0.601
## freq_CELEX_lem               -0.430
## AoA_Brysbaertetal2014         *****
## concrete_Brysbaertetal2014   <0.001
##                              concrete_Brysbaertetal2014
## letters                      -0.415
## phonemes_DUTCHPOND           -0.397
## orth_neighbours_DUTCHPOND     0.391
## phon_neighbours_DUTCHPOND     0.348
## freq_lg10CD_SUBTLEXNL         0.007
## freq_lg10WF_SUBTLEXNL         0.039
## freq_CELEX_lem               -0.124
## AoA_Brysbaertetal2014        -0.569
## concrete_Brysbaertetal2014    *****
##
## upper diagonal part contains correlation coefficient estimates
## lower diagonal part contains corresponding p-values
```

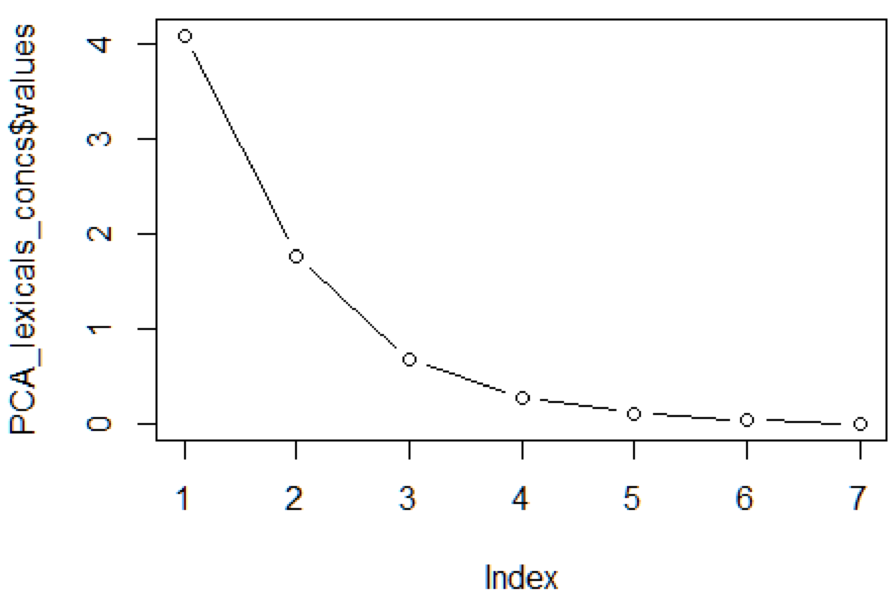
```
corrs_concs = rcor.test(mat_lexicals_concs, use='complete.obs')
write.csv(corrs_concs$cor.mat, file = "corrs_concs.csv",na="") # find table in folder


# go on to PCA. This does not include age of acquisition or concreteness for a
# better comparison with the English data, and because no correlations > .7 (i.e. half
# of variance explained)

lexicals_concs <- concs[c('letters', 'phonemes_DUTCHPOND', 'orth_neighbours_DUTCHPOND',
'phon_neighbours_DUTCHPOND', 'freq_lg10CD_SUBTLEXNL', 'freq_lg10WF_SUBTLEXNL',
'freq_CELEX_lem')]

nrow(lexicals_concs)
```

```
## [1] 411
```

```r
# start with PCA for lexical variables, done as in Lynott and Connell (2013)
# Check conditions for a PCA
# Correlations

cor(lexicals_concs, use = 'complete.obs')

##                              letters phonemes_DUTCHPOND
## letters                    1.0000000          0.9458316
## phonemes_DUTCHPOND         0.9458316          1.0000000
## orth_neighbours_DUTCHPOND -0.6300830         -0.6067762
## phon_neighbours_DUTCHPOND -0.6130258         -0.6214952
## freq_lg10CD_SUBTLEXNL     -0.3777808         -0.3760751
## freq_lg10WF_SUBTLEXNL     -0.3915313         -0.3811413
## freq_CELEX_lem            -0.2145080         -0.2359925
##                           orth_neighbours_DUTCHPOND
## letters                                  -0.6300830
## phonemes_DUTCHPOND                       -0.6067762
## orth_neighbours_DUTCHPOND                 1.0000000
## phon_neighbours_DUTCHPOND                 0.8793924
## freq_lg10CD_SUBTLEXNL                     0.3333547
## freq_lg10WF_SUBTLEXNL                     0.3431698
## freq_CELEX_lem                            0.2009497
##                           phon_neighbours_DUTCHPOND freq_lg10CD_SUBTLEXNL
## letters                                  -0.6130258            -0.3777808
## phonemes_DUTCHPOND                       -0.6214952            -0.3760751
## orth_neighbours_DUTCHPOND                 0.8793924             0.3333547
## phon_neighbours_DUTCHPOND                 1.0000000             0.3535024
## freq_lg10CD_SUBTLEXNL                     0.3535024             1.0000000
## freq_lg10WF_SUBTLEXNL                     0.3561152             0.9874908
## freq_CELEX_lem                            0.2178926             0.7693247
##                           freq_lg10WF_SUBTLEXNL freq_CELEX_lem
## letters                              -0.3915313     -0.2145080
## phonemes_DUTCHPOND                   -0.3811413     -0.2359925
## orth_neighbours_DUTCHPOND             0.3431698      0.2009497
## phon_neighbours_DUTCHPOND             0.3561152      0.2178926
## freq_lg10CD_SUBTLEXNL                 0.9874908      0.7693247
## freq_lg10WF_SUBTLEXNL                 1.0000000      0.7508591
## freq_CELEX_lem                        0.7508591      1.0000000

# Result: all variables fit for PCA, as they have few scores below .3
# The correlations broadly replicate Lynott and Connell.

# now on the raw vars:
cortest.bartlett(lexicals_concs)

## R was not square, finding R from data

## $chisq
## [1] 3798.943
##
## $p.value
```

```
## [1] 0
##
## $df
## [1] 21

# GOOD: Bartlett's test significant

# KMO: Kaiser-Meyer-Olkin Measure of Sampling Adequacy
lexicals_concs_matrix <- cor(lexicals_concs, use = 'complete.obs')
KMO(lexicals_concs_matrix)

## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = lexicals_concs_matrix)
## Overall MSA =  0.71
## MSA for each item =
##                   letters        phonemes_DUTCHPOND
##                      0.68                      0.68
## orth_neighbours_DUTCHPOND phon_neighbours_DUTCHPOND
##                      0.71                      0.72
##     freq_lg10CD_SUBTLEXNL     freq_lg10WF_SUBTLEXNL
##                      0.67                      0.68
##            freq_CELEX_lem
##                      0.94

# Result: .71 = good.

# determinant
det(lexicals_concs_matrix)

## [1] 0.000105064

# GOOD: above 0.00001

# start off with unrotated PCA

PCA_lexicals_concs <- psych::principal(lexicals_concs, nfactors = 7, scores = TRUE)
PCA_lexicals_concs

## Principal Components Analysis
## Call: psych::principal(r = lexicals_concs, nfactors = 7, scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##                             RC2   RC1   RC3   RC4   RC5   RC6   RC7 h2
## letters                   -0.19  0.91 -0.32 -0.04  0.03  0.16  0.00  1
## phonemes_DUTCHPOND        -0.16  0.92 -0.31 -0.05 -0.03 -0.16  0.00  1
## orth_neighbours_DUTCHPOND  0.15 -0.33  0.90  0.04 -0.24 -0.01  0.00  1
## phon_neighbours_DUTCHPOND  0.17 -0.33  0.90  0.05  0.24  0.01  0.00  1
## freq_lg10CD_SUBTLEXNL      0.96 -0.17  0.15  0.17  0.01  0.01  0.07  1
## freq_lg10WF_SUBTLEXNL      0.96 -0.18  0.16  0.15  0.00 -0.01 -0.07  1
## freq_CELEX_lem             0.65 -0.07  0.07  0.75  0.00  0.00  0.00  1
##                               u2 com
## letters                  4.4e-16 1.4
```

```
## phonemes_DUTCHPOND        8.9e-16 1.4
## orth_neighbours_DUTCHPOND 1.2e-15 1.5
## phon_neighbours_DUTCHPOND 2.0e-15 1.5
## freq_lg10CD_SUBTLEXNL     1.8e-15 1.2
## freq_lg10WF_SUBTLEXNL     1.6e-15 1.2
## freq_CELEX_lem            2.0e-15 2.0
##
##                         RC2  RC1  RC3  RC4  RC5  RC6  RC7
## SS loadings            2.37 1.96 1.86 0.62 0.12 0.05 0.01
## Proportion Var         0.34 0.28 0.27 0.09 0.02 0.01 0.00
## Cumulative Var         0.34 0.62 0.89 0.97 0.99 1.00 1.00
## Proportion Explained   0.34 0.28 0.27 0.09 0.02 0.01 0.00
## Cumulative Proportion  0.34 0.62 0.89 0.97 0.99 1.00 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 7 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0
##  with the empirical chi square  0  with prob <  NA
##
## Fit based upon off diagonal values = 1

# by Kaiser's and Joliffe's standard, extract 3 RCs

# scree analysis
plot(PCA_lexicals_concs$values, type = "b")
```



```r
# result: again, extract 3 components

PCA_lexicals_concs <- psych::principal(lexicals_concs, nfactors = 3, rotate =
"varimax", scores = TRUE)

PCA_lexicals_concs #-> check explained variance along components

## Principal Components Analysis
## Call: psych::principal(r = lexicals_concs, nfactors = 3, rotate = "varimax",
##     scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##                             RC2   RC1   RC3   h2    u2 com
## letters                   -0.19  0.91 -0.33 0.97 0.027 1.4
## phonemes_DUTCHPOND        -0.16  0.92 -0.33 0.97 0.027 1.3
## orth_neighbours_DUTCHPOND  0.15 -0.32  0.90 0.94 0.058 1.3
## phon_neighbours_DUTCHPOND  0.17 -0.32  0.90 0.94 0.059 1.3
## freq_lg10CD_SUBTLEXNL      0.95 -0.18  0.16 0.95 0.048 1.1
## freq_lg10WF_SUBTLEXNL      0.94 -0.19  0.17 0.94 0.056 1.2
## freq_CELEX_lem             0.89 -0.05  0.07 0.81 0.193 1.0
##
##                        RC2  RC1  RC3
## SS loadings           2.68 1.94 1.90
## Proportion Var        0.38 0.28 0.27
## Cumulative Var        0.38 0.66 0.93
## Proportion Explained  0.41 0.30 0.29
## Cumulative Proportion 0.41 0.71 1.00
##
## Mean item complexity =  1.2
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.03
##  with the empirical chi square  20.18  with prob <  0.00016
##
## Fit based upon off diagonal values = 1

PCA_lexicals_concs$loadings

##
## Loadings:
##                           RC2    RC1    RC3
## letters                   -0.185  0.910 -0.334
## phonemes_DUTCHPOND        -0.163  0.917 -0.326
## orth_neighbours_DUTCHPOND  0.148 -0.322  0.903
## phon_neighbours_DUTCHPOND  0.169 -0.317  0.901
## freq_lg10CD_SUBTLEXNL      0.945 -0.181  0.161
## freq_lg10WF_SUBTLEXNL      0.937 -0.192  0.168
## freq_CELEX_lem             0.894
##
##                  RC2   RC1   RC3
## SS loadings    2.683 1.945 1.905
```

```
## Proportion Var 0.383 0.278 0.272
## Cumulative Var 0.383 0.661 0.933

# The PCA replicates Lynott and Connell. Standdized correlation coefficients
# between each PC and its corresponding set of variables are all above .89,
# while the rest of coefficients are all below .33.

PCA_lexicals_concs

## Principal Components Analysis
## Call: psych::principal(r = lexicals_concs, nfactors = 3, rotate = "varimax",
##     scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##                              RC2   RC1   RC3   h2    u2 com
## letters                    -0.19  0.91 -0.33 0.97 0.027 1.4
## phonemes_DUTCHPOND         -0.16  0.92 -0.33 0.97 0.027 1.3
## orth_neighbours_DUTCHPOND   0.15 -0.32  0.90 0.94 0.058 1.3
## phon_neighbours_DUTCHPOND   0.17 -0.32  0.90 0.94 0.059 1.3
## freq_lg10CD_SUBTLEXNL       0.95 -0.18  0.16 0.95 0.048 1.1
## freq_lg10WF_SUBTLEXNL       0.94 -0.19  0.17 0.94 0.056 1.2
## freq_CELEX_lem              0.89 -0.05  0.07 0.81 0.193 1.0
##
##                        RC2  RC1  RC3
## SS loadings           2.68 1.94 1.90
## Proportion Var        0.38 0.28 0.27
## Cumulative Var        0.38 0.66 0.93
## Proportion Explained  0.41 0.30 0.29
## Cumulative Proportion 0.41 0.71 1.00
##
## Mean item complexity =  1.2
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.03
##  with the empirical chi square  20.18  with prob <  0.00016
##
## Fit based upon off diagonal values = 1

# RC1 = length // RC2 = frequency // RC3 = distinctiveness

PCA_lexicals_concs$residual

##                                letters phonemes_DUTCHPOND
## letters                    0.026594010       -0.025634342
## phonemes_DUTCHPOND        -0.025634342        0.027463542
## orth_neighbours_DUTCHPOND -0.007939846        0.007595230
## phon_neighbours_DUTCHPOND  0.007967311       -0.008984482
## freq_lg10CD_SUBTLEXNL      0.004139691        0.006696974
## freq_lg10WF_SUBTLEXNL      0.001306927        0.010591926
## freq_CELEX_lem            -0.005090237       -0.018061566
##                           orth_neighbours_DUTCHPOND
## letters                                -0.007939846
```

```
## phonemes_DUTCHPOND                          0.007595230
## orth_neighbours_DUTCHPOND                   0.058173584
## phon_neighbours_DUTCHPOND                  -0.058160242
## freq_lg10CD_SUBTLEXNL                      -0.003609941
## freq_lg10WF_SUBTLEXNL                      -0.000955949
## freq_CELEX_lem                              0.005890390
##                           phon_neighbours_DUTCHPOND freq_lg10CD_SUBTLEXNL
## letters                                     0.007967311           0.004139691
## phonemes_DUTCHPOND                         -0.008984482           0.006696974
## orth_neighbours_DUTCHPOND                  -0.058160242          -0.003609941
## phon_neighbours_DUTCHPOND                   0.058821233          -0.001681398
## freq_lg10CD_SUBTLEXNL                      -0.001681398           0.048039445
## freq_lg10WF_SUBTLEXNL                      -0.004899638           0.040904630
## freq_CELEX_lem                              0.005469160          -0.090683850
##                           freq_lg10WF_SUBTLEXNL freq_CELEX_lem
## letters                                 0.001306927   -0.005090237
## phonemes_DUTCHPOND                      0.010591926   -0.018061566
## orth_neighbours_DUTCHPOND              -0.000955949    0.005890390
## phon_neighbours_DUTCHPOND              -0.004899638    0.005469160
## freq_lg10CD_SUBTLEXNL                   0.040904630   -0.090683850
## freq_lg10WF_SUBTLEXNL                   0.055871031   -0.098544553
## freq_CELEX_lem                         -0.098544553    0.192845909

PCA_lexicals_concs$fit

## [1] 0.9950821

PCA_lexicals_concs$communality

##                   letters        phonemes_DUTCHPOND
##                 0.9734060                 0.9725365
## orth_neighbours_DUTCHPOND phon_neighbours_DUTCHPOND
##                 0.9418264                 0.9411788
##     freq_lg10CD_SUBTLEXNL     freq_lg10WF_SUBTLEXNL
##                 0.9519606                 0.9441290
##            freq_CELEX_lem
##                 0.8071541

# Results based on a Kaiser-normalizalized orthogonal (varimax) rotation
# (by default in psych::stats pack). Residuals good: less than half w/ absolute
# values > 0.05. Model fit good, > .90. Communalities (h2) good, all well > .7

concs <- cbind(concs, PCA_lexicals_concs$scores)


# REGRESSION

# standardize (mean-center and scale)
concs$s_Auditory <- scale(concs$Auditory)
concs$s_Haptic <- scale(concs$Haptic)
concs$s_Visual <- scale(concs$Visual)
```

```
concs$s_freq_lg10CD_SUBTLEXNL <- scale(concs$freq_lg10CD_SUBTLEXNL)
concs$s_freq_lg10WF_SUBTLEXNL <- scale(concs$freq_lg10WF_SUBTLEXNL)
concs$s_freq_CELEX_lem <- scale(concs$freq_CELEX_lem)
concs$s_AoA_Brysbaertetal2014 <- scale(concs$AoA_Brysbaertetal2014)
concs$s_concrete_Brysbaertetal2014 <- scale(concs$concrete_Brysbaertetal2014)
concs$s_letters <- scale(concs$letters)
concs$s_phonemes_DUTCHPOND <- scale(concs$phonemes_DUTCHPOND)
concs$s_orth_neighbours_DUTCHPOND <- scale(concs$orth_neighbours_DUTCHPOND)
concs$s_phon_neighbours_DUTCHPOND <- scale(concs$phon_neighbours_DUTCHPOND)
concs$s_RC1_lexicals <- scale(concs$RC1)
concs$s_RC2_lexicals <- scale(concs$RC2)
concs$s_RC3_lexicals <- scale(concs$RC3)

# length: letters
fit_letters_concs <- lm(concs$s_letters ~ concs$s_Auditory + concs$s_Haptic +
concs$s_Visual, data = concs)
stat.desc(fit_letters_concs$residuals, norm = TRUE)

##                          x
## nbr.val       4.110000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.869411e+00
## max           4.100072e+00
## range         5.969482e+00
## sum          -2.588207e-15
## median       -1.709015e-01
## mean         -6.333218e-18
## SE.mean       4.751817e-02
## CI.mean.0.95  9.340965e-02
## var           9.280284e-01
## std.dev       9.633423e-01
## coef.var     -1.521095e+17
## skewness      7.313500e-01
## skew.2SE      3.037507e+00
## kurtosis      3.972703e-01
## kurt.2SE      8.269597e-01
## normtest.W    9.582384e-01
## normtest.p    2.101496e-09

# residuals distribution: skew. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(concs$s_letters)

##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 411    0  1  -0.28   -0.09 1.17 -1.46 4.05  5.51 0.74     0.28
##      se
## X1 0.05
```

```r
concs$log_s_letters <- log(3 + concs$s_letters)

fit_letters_concs <- lm(concs$log_s_letters ~ concs$s_Auditory + concs$s_Haptic +
concs$s_Visual, data = concs)

# check residuals again
stat.desc(fit_letters_concs$residuals, norm = TRUE)

##                          x
## nbr.val       4.110000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -7.546621e-01
## max           9.308451e-01
## range         1.685507e+00
## sum           9.228729e-16
## median       -1.147895e-02
## mean          2.265877e-18
## SE.mean       1.549036e-02
## CI.mean.0.95  3.045044e-02
## var           9.862002e-02
## std.dev       3.140382e-01
## coef.var      1.385946e+17
## skewness      9.384492e-02
## skew.2SE      3.897649e-01
## kurtosis     -6.833286e-01
## kurt.2SE     -1.422420e+00
## normtest.W    9.883934e-01
## normtest.p    2.360257e-03

# better though still skew/kurtose

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1),
# and tolerance (pref. > 0.2)
vif(fit_letters_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##          1.009925         1.245338         1.254282

mean(vif(fit_letters_concs))

## [1] 1.169848

1/vif(fit_letters_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##         0.9901729        0.8029946        0.7972690

# RESULTS: all good

step_letters_concs_AIC <- stepAIC(fit_letters_concs, direction="both")
```

```
## Start:  AIC=-945.07
## concs$log_s_letters ~ concs$s_Auditory + concs$s_Haptic + concs$s_Visual
## 
##                      Df Sum of Sq    RSS     AIC
## - concs$s_Visual    1   0.19201 40.626 -945.13
## <none>                          40.434 -945.07
## - concs$s_Auditory  1   0.99968 41.434 -937.04
## - concs$s_Haptic    1   1.51007 41.944 -932.01
## 
## Step:  AIC=-945.13
## concs$log_s_letters ~ concs$s_Auditory + concs$s_Haptic
## 
##                      Df Sum of Sq    RSS     AIC
## <none>                          40.626 -945.13
## + concs$s_Visual    1   0.19201 40.434 -945.07
## - concs$s_Auditory  1   0.92402 41.550 -937.88
## - concs$s_Haptic    1   2.52263 43.149 -922.37
```

```
step_letters_concs_F <- stepAIC(fit_letters_concs, direction="both", test="F")
```

```
## Start:  AIC=-945.07
## concs$log_s_letters ~ concs$s_Auditory + concs$s_Haptic + concs$s_Visual
## 
##                      Df Sum of Sq    RSS     AIC F Value     Pr(F)
## - concs$s_Visual    1   0.19201 40.626 -945.13  1.9327 0.1652242
## <none>                          40.434 -945.07
## - concs$s_Auditory  1   0.99968 41.434 -937.04 10.0625 0.0016279 **
## - concs$s_Haptic    1   1.51007 41.944 -932.01 15.1999 0.0001131 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Step:  AIC=-945.13
## concs$log_s_letters ~ concs$s_Auditory + concs$s_Haptic
## 
##                      Df Sum of Sq    RSS     AIC F Value     Pr(F)
## <none>                          40.626 -945.13
## + concs$s_Visual    1   0.19201 40.434 -945.07  1.9327  0.165224
## - concs$s_Auditory  1   0.92402 41.550 -937.88  9.2797  0.002468 **
## - concs$s_Haptic    1   2.52263 43.149 -922.37 25.3342 7.245e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_letters_concs)
```

```
## 
## Call:
## lm(formula = concs$log_s_letters ~ concs$s_Auditory + concs$s_Haptic + 
##     concs$s_Visual, data = concs)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.75466 -0.24161 -0.01148  0.23692  0.93085
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.04489    0.01555  67.207  < 2e-16 ***
## concs$s_Auditory  0.04962    0.01564   3.172 0.001628 **
## concs$s_Haptic   -0.06773    0.01737  -3.899 0.000113 ***
## concs$s_Visual   -0.02424    0.01743  -1.390 0.165224
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3152 on 407 degrees of freedom
## Multiple R-squared:  0.08311,    Adjusted R-squared:  0.07635
## F-statistic:  12.3 on 3 and 407 DF,  p-value: 1.025e-07
```

```
# length: phonemes_DUTCHPOND
fit_phonemes_DUTCHPOND_concs <- lm(concs$s_phonemes_DUTCHPOND ~ concs$s_Auditory +
concs$s_Haptic + concs$s_Visual, data = concs)
stat.desc(fit_phonemes_DUTCHPOND_concs$residuals, norm = TRUE)
```

```
##                          x
## nbr.val       3.890000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -2.130401e+00
## max           4.299017e+00
## range         6.429418e+00
## sum           5.016820e-15
## median       -1.585982e-01
## mean          1.286522e-17
## SE.mean       4.897103e-02
## CI.mean.0.95  9.628179e-02
## var           9.328850e-01
## std.dev       9.658597e-01
## coef.var      7.507528e+16
## skewness      6.756933e-01
## skew.2SE      2.730764e+00
## kurtosis      6.167434e-01
## kurt.2SE      1.249403e+00
## normtest.W    9.707429e-01
## normtest.p    4.832522e-07
```

```
# residuals distribution: skew and kurtose. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(concs$s_phonemes_DUTCHPOND)
```

```
##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 389    0  1   0.08   -0.07 1.36 -1.76 4.21  5.97  0.7     0.41
##      se
## X1 0.05
```

```r
concs$log_s_phonemes_DUTCHPOND <- log(3 + concs$s_phonemes_DUTCHPOND)

fit_phonemes_DUTCHPOND_concs <- lm(concs$log_s_phonemes_DUTCHPOND ~ concs$s_Auditory
 + concs$s_Haptic + concs$s_Visual, data = concs)

# check residuals again
stat.desc(fit_phonemes_DUTCHPOND_concs$residuals, norm = TRUE)

##                           x
## nbr.val       3.890000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -9.530293e-01
## max           9.618234e-01
## range         1.914853e+00
## sum          -4.623038e-15
## median       -2.033824e-03
## mean         -1.187383e-17
## SE.mean       1.628015e-02
## CI.mean.0.95  3.200835e-02
## var           1.031018e-01
## std.dev       3.210947e-01
## coef.var     -2.704223e+16
## skewness     -1.089115e-01
## skew.2SE     -4.401577e-01
## kurtosis     -2.710183e-01
## kurt.2SE     -5.490309e-01
## normtest.W    9.942364e-01
## normtest.p    1.496887e-01

# good

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_phonemes_DUTCHPOND_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##          1.005252         1.250064         1.252494

mean(vif(fit_phonemes_DUTCHPOND_concs))

## [1] 1.16927

1/vif(fit_phonemes_DUTCHPOND_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##         0.9947753        0.7999588        0.7984071

# RESULTS: all good
```

```r
step_phonemes_DUTCHPOND_concs_AIC <- stepAIC(fit_phonemes_DUTCHPOND_concs,
direction="both")
```

```
## Start:  AIC=-876.82
## concs$log_s_phonemes_DUTCHPOND ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                     Df Sum of Sq    RSS     AIC
## - concs$s_Visual    1   0.07544 40.079 -878.09
## <none>                          40.004 -876.82
## - concs$s_Haptic    1   1.11568 41.119 -868.12
## - concs$s_Auditory  1   1.20234 41.206 -867.30
##
## Step:  AIC=-878.09
## concs$log_s_phonemes_DUTCHPOND ~ concs$s_Auditory + concs$s_Haptic
##
##                     Df Sum of Sq    RSS     AIC
## <none>                          40.079 -878.09
## + concs$s_Visual    1   0.07544 40.004 -876.82
## - concs$s_Auditory  1   1.16707 41.246 -868.93
## - concs$s_Haptic    1   1.73650 41.815 -863.59
```

```r
step_phonemes_DUTCHPOND_concs_F <- stepAIC(fit_phonemes_DUTCHPOND_concs,
direction="both", test="F")
```

```
## Start:  AIC=-876.82
## concs$log_s_phonemes_DUTCHPOND ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                     Df Sum of Sq    RSS     AIC F Value     Pr(F)
## - concs$s_Visual    1   0.07544 40.079 -878.09  0.7261 0.3946957
## <none>                          40.004 -876.82
## - concs$s_Haptic    1   1.11568 41.119 -868.12 10.7374 0.0011452 **
## - concs$s_Auditory  1   1.20234 41.206 -867.30 11.5715 0.0007399 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-878.09
## concs$log_s_phonemes_DUTCHPOND ~ concs$s_Auditory + concs$s_Haptic
##
##                     Df Sum of Sq    RSS     AIC F Value     Pr(F)
## <none>                          40.079 -878.09
## + concs$s_Visual    1   0.07544 40.004 -876.82  0.7261 0.3946957
## - concs$s_Auditory  1   1.16707 41.246 -868.93 11.2401 0.0008796 ***
## - concs$s_Haptic    1   1.73650 41.815 -863.59 16.7242 5.263e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(fit_phonemes_DUTCHPOND_concs)
```

```
##
## Call:
## lm(formula = concs$log_s_phonemes_DUTCHPOND ~ concs$s_Auditory +
##     concs$s_Haptic + concs$s_Visual, data = concs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95303 -0.22895 -0.00203  0.23763  0.96182
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.04396    0.01636  63.821  < 2e-16 ***
## concs$s_Auditory  0.05577    0.01640   3.402  0.00074 ***
## concs$s_Haptic   -0.05941    0.01813  -3.277  0.00115 **
## concs$s_Visual   -0.01571    0.01844  -0.852  0.39470
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3223 on 385 degrees of freedom
##   (22 observations deleted due to missingness)
## Multiple R-squared:  0.07083,    Adjusted R-squared:  0.06359
## F-statistic: 9.783 on 3 and 385 DF,  p-value: 3.109e-06
```

```
# distinctiveness: orth neigh size
fit_orth_neighbours_DUTCHPOND_concs <- lm(concs$s_orth_neighbours_DUTCHPOND ~
concs$s_Auditory + concs$s_Haptic + concs$s_Visual, data = concs)
stat.desc(fit_orth_neighbours_DUTCHPOND_concs$residuals, norm = TRUE)
```

```
##                          x
## nbr.val       4.110000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.418943e+00
## max           4.983802e+00
## range         6.402745e+00
## sum           1.929013e-15
## median       -3.388470e-01
## mean          4.732767e-18
## SE.mean       4.699016e-02
## CI.mean.0.95  9.237170e-02
## var           9.075189e-01
## std.dev       9.526379e-01
## coef.var      2.012856e+17
## skewness      1.758073e+00
## skew.2SE      7.301783e+00
## kurtosis      3.397510e+00
## kurt.2SE      7.072272e+00
## normtest.W    8.209900e-01
## normtest.p    4.657007e-21
```

```r
# residuals distribution: skewed and kurtosed. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(concs$s_orth_neighbours_DUTCHPOND)

##    vars   n mean sd median trimmed  mad  min  max range skew kurtosis   se
## X1    1 411    0  1  -0.53   -0.22 0.26 -0.7 4.88  5.59 1.88     3.42 0.05

concs$log_s_orth_neighbours_DUTCHPOND <- log(2 + concs$s_orth_neighbours_DUTCHPOND)

fit_orth_neighbours_DUTCHPOND_concs <- lm(concs$log_s_orth_neighbours_DUTCHPOND ~
concs$s_Auditory + concs$s_Haptic + concs$s_Visual, data = concs)

# check residuals again
stat.desc(fit_orth_neighbours_DUTCHPOND_concs$residuals, norm = TRUE)

##                          x
## nbr.val       4.110000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -6.413729e-01
## max           1.377855e+00
## range         2.019228e+00
## sum           2.515349e-15
## median       -1.365831e-01
## mean          6.152518e-18
## SE.mean       1.874673e-02
## CI.mean.0.95  3.685169e-02
## var           1.444417e-01
## std.dev       3.800549e-01
## coef.var      6.177226e+16
## skewness      1.033319e+00
## skew.2SE      4.291669e+00
## kurtosis      3.381673e-01
## kurt.2SE      7.039305e-01
## normtest.W    9.016737e-01
## normtest.p    1.225133e-15

# better though still skew/kurtose

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_orth_neighbours_DUTCHPOND_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##         1.009925         1.245338         1.254282

mean(vif(fit_orth_neighbours_DUTCHPOND_concs))

## [1] 1.169848
```

```r
1/vif(fit_orth_neighbours_DUTCHPOND_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##         0.9901729        0.8029946        0.7972690

# RESULTS: all good

step_orth_neighbours_DUTCHPOND_concs_AIC <-
stepAIC(fit_orth_neighbours_DUTCHPOND_concs, direction="both")

## Start:  AIC=-788.24
## concs$log_s_orth_neighbours_DUTCHPOND ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                      Df Sum of Sq    RSS     AIC
## - concs$s_Visual    1    0.0959 59.317 -789.57
## <none>                          59.221 -788.24
## - concs$s_Auditory  1    0.9578 60.179 -783.64
## - concs$s_Haptic    1    3.8929 63.114 -764.07
##
## Step:  AIC=-789.57
## concs$log_s_orth_neighbours_DUTCHPOND ~ concs$s_Auditory + concs$s_Haptic
##
##                      Df Sum of Sq    RSS     AIC
## <none>                          59.317 -789.57
## + concs$s_Visual    1    0.0959 59.221 -788.24
## - concs$s_Auditory  1    0.9077 60.225 -785.33
## - concs$s_Haptic    1    5.5464 64.863 -754.83

step_orth_neighbours_DUTCHPOND_concs_F <-
stepAIC(fit_orth_neighbours_DUTCHPOND_concs, direction="both", test="F")

## Start:  AIC=-788.24
## concs$log_s_orth_neighbours_DUTCHPOND ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value     Pr(F)
## - concs$s_Visual    1    0.0959 59.317 -789.57  0.6590   0.41737
## <none>                          59.221 -788.24
## - concs$s_Auditory  1    0.9578 60.179 -783.64  6.5826   0.01065 *
## - concs$s_Haptic    1    3.8929 63.114 -764.07 26.7539 3.634e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-789.57
## concs$log_s_orth_neighbours_DUTCHPOND ~ concs$s_Auditory + concs$s_Haptic
##
##                      Df Sum of Sq    RSS     AIC F Value     Pr(F)
## <none>                          59.317 -789.57
## + concs$s_Visual    1    0.0959 59.221 -788.24   0.659   0.41737
## - concs$s_Auditory  1    0.9077 60.225 -785.33   6.244   0.01286 *
```

```
## - concs$s_Haptic     1    5.5464 64.863 -754.83  38.150 1.586e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_orth_neighbours_DUTCHPOND_concs)
```

```
##
## Call:
## lm(formula = concs$log_s_orth_neighbours_DUTCHPOND ~ concs$s_Auditory +
##     concs$s_Haptic + concs$s_Visual, data = concs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6414 -0.2734 -0.1366  0.1982  1.3779
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.60117    0.01882  31.951  < 2e-16 ***
## concs$s_Auditory -0.04857    0.01893  -2.566   0.0107 *
## concs$s_Haptic    0.10874    0.02102   5.172 3.63e-07 ***
## concs$s_Visual    0.01713    0.02110   0.812   0.4174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3815 on 407 degrees of freedom
## Multiple R-squared:  0.1001, Adjusted R-squared:  0.09348
## F-statistic: 15.09 on 3 and 407 DF,  p-value: 2.477e-09
```

```
# distinctiveness: phon neigh size
fit_phon_neighbours_DUTCHPOND_concs <- lm(concs$s_phon_neighbours_DUTCHPOND ~
concs$s_Auditory + concs$s_Haptic + concs$s_Visual, data = concs)
stat.desc(fit_phon_neighbours_DUTCHPOND_concs$residuals, norm = TRUE)
```

```
##                          x
## nbr.val       4.110000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.352384e+00
## max           4.807540e+00
## range         6.159924e+00
## sum          -1.264613e-15
## median       -3.437850e-01
## mean         -3.056870e-18
## SE.mean       4.739867e-02
## CI.mean.0.95  9.317474e-02
## var           9.233666e-01
## std.dev       9.609197e-01
## coef.var     -3.143476e+17
## skewness      2.025920e+00
## skew.2SE      8.414228e+00
## kurtosis      4.788769e+00
```

```
## kurt.2SE      9.968323e+00
## normtest.W    7.877461e-01
## normtest.p    9.278171e-23

# residuals distribution: skewed and kurtosed. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(concs$s_phon_neighbours_DUTCHPOND)

##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 411    0  1  -0.55   -0.22 0.18 -0.67 5.26  5.92 2.06     4.49
##      se
## X1 0.05

concs$log_s_phon_neighbours_DUTCHPOND <- log(2 + concs$s_phon_neighbours_DUTCHPOND)

fit_phon_neighbours_DUTCHPOND_concs <- lm(concs$log_s_phon_neighbours_DUTCHPOND ~
concs$s_Auditory + concs$s_Haptic + concs$s_Visual, data = concs)

# check residuals again
stat.desc(fit_phon_neighbours_DUTCHPOND_concs$residuals, norm = TRUE)

##                          x
## nbr.val       4.110000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -6.105674e-01
## max           1.390839e+00
## range         2.001407e+00
## sum           2.714842e-15
## median       -1.406069e-01
## mean          6.635001e-18
## SE.mean       1.839574e-02
## CI.mean.0.95  3.616174e-02
## var           1.390838e-01
## std.dev       3.729394e-01
## coef.var      5.620788e+16
## skewness      1.199464e+00
## skew.2SE      4.981718e+00
## kurtosis      9.001905e-01
## kurt.2SE      1.873840e+00
## normtest.W    8.836707e-01
## normtest.p    4.522529e-17

# better but not perfect

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_phon_neighbours_DUTCHPOND_concs)
```

```
## concs$s_Auditory   concs$s_Haptic   concs$s_Visual 
##         1.009925         1.245338         1.254282

mean(vif(fit_phon_neighbours_DUTCHPOND_concs))

## [1] 1.169848

1/vif(fit_phon_neighbours_DUTCHPOND_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual 
##        0.9901729        0.8029946        0.7972690

# RESULTS: all good

step_phon_neighbours_DUTCHPOND_concs_AIC <-
stepAIC(fit_phon_neighbours_DUTCHPOND_concs, direction="both")

## Start:  AIC=-803.77
## concs$log_s_phon_neighbours_DUTCHPOND ~ concs$s_Auditory + concs$s_Haptic + 
##     concs$s_Visual
## 
##                    Df Sum of Sq    RSS     AIC
## - concs$s_Visual    1    0.0168 57.041 -805.65
## <none>                          57.024 -803.77
## - concs$s_Auditory  1    0.6885 57.713 -800.84
## - concs$s_Haptic    1    3.8278 60.852 -779.07
## 
## Step:  AIC=-805.65
## concs$log_s_phon_neighbours_DUTCHPOND ~ concs$s_Auditory + concs$s_Haptic
## 
##                    Df Sum of Sq    RSS     AIC
## <none>                          57.041 -805.65
## + concs$s_Visual    1    0.0168 57.024 -803.77
## - concs$s_Auditory  1    0.6740 57.715 -802.82
## - concs$s_Haptic    1    5.0509 62.092 -772.78

step_phon_neighbours_DUTCHPOND_concs_F <- stepAIC(fit_phon_neighbours_DUTCHPOND_concs,
direction="both", test="F")

## Start:  AIC=-803.77
## concs$log_s_phon_neighbours_DUTCHPOND ~ concs$s_Auditory + concs$s_Haptic + 
##     concs$s_Visual
## 
##                    Df Sum of Sq    RSS     AIC F Value     Pr(F)    
## - concs$s_Visual    1    0.0168 57.041 -805.65  0.1199   0.72937    
## <none>                          57.024 -803.77                      
## - concs$s_Auditory  1    0.6885 57.713 -800.84  4.9142   0.02719 *  
## - concs$s_Haptic    1    3.8278 60.852 -779.07 27.3204 2.761e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Step:  AIC=-805.65
```

```
## concs$log_s_phon_neighbours_DUTCHPOND ~ concs$s_Auditory + concs$s_Haptic
##
##                      Df Sum of Sq    RSS     AIC F Value     Pr(F)
## <none>                           57.041 -805.65
## + concs$s_Visual      1    0.0168 57.024 -803.77   0.120   0.72937
## - concs$s_Auditory    1    0.6740 57.715 -802.82   4.821   0.02868 *
## - concs$s_Haptic      1    5.0509 62.092 -772.78  36.128 4.101e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_phon_neighbours_DUTCHPOND_concs)
```

```
##
## Call:
## lm(formula = concs$log_s_phon_neighbours_DUTCHPOND ~ concs$s_Auditory +
##     concs$s_Haptic + concs$s_Visual, data = concs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6106 -0.2524 -0.1406  0.1897  1.3908
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.604084   0.018463  32.718  < 2e-16 ***
## concs$s_Auditory -0.041182   0.018577  -2.217   0.0272 *
## concs$s_Haptic    0.107827   0.020629   5.227 2.76e-07 ***
## concs$s_Visual    0.007167   0.020703   0.346   0.7294
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3743 on 407 degrees of freedom
## Multiple R-squared:  0.09194,    Adjusted R-squared:  0.08525
## F-statistic: 13.74 on 3 and 407 DF,  p-value: 1.499e-08
```

```
# freq: SUBTLEX-NL log-10 CD

fit_freq_lg10CD_SUBTLEXNL_concs <- lm(concs$s_freq_lg10CD_SUBTLEXNL ~
concs$s_Auditory + concs$s_Haptic + concs$s_Visual, data = concs)
stat.desc(fit_freq_lg10CD_SUBTLEXNL_concs$residuals, norm = TRUE)
```

```
##                          x
## nbr.val       4.070000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -3.048064e+00
## max           2.307864e+00
## range         5.355928e+00
## sum          -7.049916e-15
## median        1.144372e-02
## mean         -1.729615e-17
## SE.mean       4.797448e-02
```

```
## CI.mean.0.95  9.430939e-02
## var           9.367312e-01
## std.dev       9.678487e-01
## coef.var     -5.595745e+16
## skewness     -2.807029e-01
## skew.2SE     -1.160194e+00
## kurtosis      7.175888e-02
## kurt.2SE      1.486536e-01
## normtest.W    9.927886e-01
## normtest.p    4.745518e-02

# residuals distribution: skew. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(concs$s_freq_lg10CD_SUBTLEXNL)

##    vars   n mean sd median trimmed  mad   min  max range  skew kurtosis
## X1    1 407    0  1   0.07    0.04 1.01 -3.11 1.88  4.99 -0.39    -0.14
##      se
## X1 0.05

concs$log_s_freq_lg10CD_SUBTLEXNL <- log(5 + concs$s_freq_lg10CD_SUBTLEXNL)

fit_freq_lg10CD_SUBTLEXNL_concs <- lm(concs$log_s_freq_lg10CD_SUBTLEXNL ~
concs$s_Auditory + concs$s_Haptic + concs$s_Visual, data = concs)

# check residuals again
stat.desc(fit_freq_lg10CD_SUBTLEXNL_concs$residuals, norm = TRUE)

##                           x
## nbr.val       4.070000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -9.360006e-01
## max           4.459662e-01
## range         1.381967e+00
## sum          -2.066923e-15
## median        2.387622e-02
## mean         -5.070726e-18
## SE.mean       1.060014e-02
## CI.mean.0.95  2.083802e-02
## var           4.573176e-02
## std.dev       2.138498e-01
## coef.var     -4.217342e+16
## skewness     -9.892655e-01
## skew.2SE     -4.088806e+00
## kurtosis      1.832425e+00
## kurt.2SE      3.795999e+00
## normtest.W    9.507730e-01
## normtest.p    2.130839e-10
```

```
# worse! back
fit_freq_lg10CD_SUBTLEXNL_concs <- lm(concs$s_freq_lg10CD_SUBTLEXNL ~
concs$s_Auditory + concs$s_Haptic + concs$s_Visual, data = concs)

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_freq_lg10CD_SUBTLEXNL_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##         1.007087         1.251386         1.256979

mean(vif(fit_freq_lg10CD_SUBTLEXNL_concs))

## [1] 1.171817

1/vif(fit_freq_lg10CD_SUBTLEXNL_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##        0.9929629        0.7991139        0.7955584

# RESULTS: all good

step_freq_lg10CD_SUBTLEXNL_concs_AIC <- stepAIC(fit_freq_lg10CD_SUBTLEXNL_concs,
direction="both")

## Start:  AIC=-19.6
## concs$s_freq_lg10CD_SUBTLEXNL ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                      Df Sum of Sq    RSS      AIC
## - concs$s_Haptic     1     1.122 381.43 -20.4034
## <none>                           380.31 -19.6023
## - concs$s_Visual     1     3.709 384.02 -17.6523
## - concs$s_Auditory   1    16.168 396.48  -4.6579
##
## Step:  AIC=-20.4
## concs$s_freq_lg10CD_SUBTLEXNL ~ concs$s_Auditory + concs$s_Visual
##
##                      Df Sum of Sq    RSS      AIC
## <none>                           381.43 -20.4034
## + concs$s_Haptic     1    1.1220 380.31 -19.6023
## - concs$s_Visual     1    7.2094 388.64 -14.7826
## - concs$s_Auditory   1   15.7897 397.22  -5.8947

step_freq_lg10CD_SUBTLEXNL__concsF <- stepAIC(fit_freq_lg10CD_SUBTLEXNL_concs,
direction="both", test="F")

## Start:  AIC=-19.6
## concs$s_freq_lg10CD_SUBTLEXNL ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                      Df Sum of Sq    RSS      AIC F Value     Pr(F)
```

```
## - concs$s_Haptic    1     1.122 381.43 -20.4034  1.1889    0.2762
## <none>                         380.31 -19.6023
## - concs$s_Visual    1     3.709 384.02 -17.6523  3.9302    0.0481 *
## - concs$s_Auditory  1    16.168 396.48  -4.6579 17.1321 4.248e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-20.4
## concs$s_freq_lg10CD_SUBTLEXNL ~ concs$s_Auditory + concs$s_Visual
##
##                    Df Sum of Sq    RSS      AIC F Value     Pr(F)
## <none>                          381.43 -20.4034
## + concs$s_Haptic    1    1.1220 380.31 -19.6023  1.1889  0.276198
## - concs$s_Visual    1    7.2094 388.64 -14.7826  7.6359  0.005983 **
## - concs$s_Auditory  1   15.7897 397.22  -5.8947 16.7238 5.219e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_freq_lg10CD_SUBTLEXNL_concs)
```

```
##
## Call:
## lm(formula = concs$s_freq_lg10CD_SUBTLEXNL ~ concs$s_Auditory +
##     concs$s_Haptic + concs$s_Visual, data = concs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.04806 -0.54760  0.01144  0.66048  2.30786
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.00433    0.04816  -0.090   0.9284
## concs$s_Auditory  0.20136    0.04865   4.139 4.25e-05 ***
## concs$s_Haptic    0.05869    0.05382   1.090   0.2762
## concs$s_Visual    0.10775    0.05435   1.982   0.0481 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9714 on 403 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.06327,    Adjusted R-squared:  0.0563
## F-statistic: 9.073 on 3 and 403 DF,  p-value: 7.971e-06
```

```
# freq: SUBTLEX-NL log-10 WF
fit_freq_lg10WF_SUBTLEXNL_concs <-
lm(concs$s_freq_lg10WF_SUBTLEXNL ~ concs$s_Auditory + concs$s_Haptic + concs$s_Visual,
 data = concs)
stat.desc(fit_freq_lg10WF_SUBTLEXNL_concs$residuals, norm = TRUE)
```

```
##                             x
## nbr.val       4.070000e+02
```

```
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -2.895265e+00
## max           2.942859e+00
## range         5.838124e+00
## sum           6.078471e-15
## median        2.548493e-02
## mean          1.480728e-17
## SE.mean       4.802025e-02
## CI.mean.0.95  9.439937e-02
## var           9.385194e-01
## std.dev       9.687721e-01
## coef.var      6.542537e+16
## skewness      2.645810e-02
## skew.2SE      1.093559e-01
## kurtosis      1.608904e-01
## kurt.2SE      3.332959e-01
## normtest.W    9.969441e-01
## normtest.p    6.460055e-01

# residuals distribution: good. Raw scores/2.SE < 1

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_freq_lg10WF_SUBTLEXNL_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##          1.007087         1.251386         1.256979

mean(vif(fit_freq_lg10WF_SUBTLEXNL_concs))

## [1] 1.171817

1/vif(fit_freq_lg10WF_SUBTLEXNL_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##         0.9929629        0.7991139        0.7955584

# RESULTS: all good

step_freq_lg10WF_SUBTLEXNL_concs_AIC <- stepAIC(fit_freq_lg10WF_SUBTLEXNL_concs,
direction="both")

## Start:  AIC=-18.83
## concs$s_freq_lg10WF_SUBTLEXNL ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                      Df Sum of Sq    RSS      AIC
## - concs$s_Haptic     1    0.6561 381.70 -20.1258
## <none>                           381.04 -18.8261
## - concs$s_Visual     1    4.3316 385.37 -16.2255
```

```
## - concs$s_Auditory  1   15.5396 396.58  -4.5572
##
## Step:  AIC=-20.13
## concs$s_freq_lg10WF_SUBTLEXNL ~ concs$s_Auditory + concs$s_Visual
##
##                     Df Sum of Sq    RSS      AIC
## <none>                           381.70 -20.1258
## + concs$s_Haptic    1    0.6561 381.04 -18.8261
## - concs$s_Visual    1    7.4743 389.17 -14.2330
## - concs$s_Auditory  1   15.2642 396.96  -6.1667
```

```
step_freq_lg10WF_SUBTLEXNL_concs_F <- stepAIC(fit_freq_lg10WF_SUBTLEXNL_concs,
direction="both", test="F")
```

```
## Start:  AIC=-18.83
## concs$s_freq_lg10WF_SUBTLEXNL ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                     Df Sum of Sq    RSS      AIC F Value     Pr(F)
## - concs$s_Haptic    1    0.6561 381.70 -20.1258  0.6940   0.40531
## <none>                           381.04 -18.8261
## - concs$s_Visual    1    4.3316 385.37 -16.2255  4.5812   0.03292 *
## - concs$s_Auditory  1   15.5396 396.58  -4.5572 16.4352 6.043e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-20.13
## concs$s_freq_lg10WF_SUBTLEXNL ~ concs$s_Auditory + concs$s_Visual
##
##                     Df Sum of Sq    RSS      AIC F Value     Pr(F)
## <none>                           381.70 -20.1258
## + concs$s_Haptic    1    0.6561 381.04 -18.8261  0.6940  0.405312
## - concs$s_Visual    1    7.4743 389.17 -14.2330  7.9111  0.005153 **
## - concs$s_Auditory  1   15.2642 396.96  -6.1667 16.1561 6.959e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_freq_lg10WF_SUBTLEXNL_concs)
```

```
##
## Call:
## lm(formula = concs$s_freq_lg10WF_SUBTLEXNL ~ concs$s_Auditory +
##     concs$s_Haptic + concs$s_Visual, data = concs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.89527 -0.59131  0.02548  0.59277  2.94286
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.004361   0.048207  -0.090   0.9280
```

```
## concs$s_Auditory  0.197415   0.048696   4.054 6.04e-05 ***
## concs$s_Haptic    0.044880   0.053875   0.833   0.4053
## concs$s_Visual    0.116441   0.054402   2.140   0.0329 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9724 on 403 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.06148,    Adjusted R-squared:  0.05449
## F-statistic:   8.8 on 3 and 403 DF,  p-value: 1.155e-05
```

```
# freq: CELEX log-10 lemma WF
fit_freq_CELEX_lem_concs <- lm(concs$s_freq_CELEX_lem ~ concs$s_Auditory +
concs$s_Haptic + concs$s_Visual, data = concs)
stat.desc(fit_freq_CELEX_lem_concs$residuals, norm = TRUE)
```

```
##                           x
## nbr.val       3.990000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -2.596448e+00
## max           2.747406e+00
## range         5.343854e+00
## sum          -3.178013e-15
## median        1.094298e-02
## mean         -7.790903e-18
## SE.mean       4.913264e-02
## CI.mean.0.95  9.659194e-02
## var           9.631926e-01
## std.dev       9.814237e-01
## coef.var     -1.259705e+17
## skewness     -3.551406e-02
## skew.2SE     -1.453467e-01
## kurtosis     -8.965432e-02
## kurt.2SE     -1.839131e-01
## normtest.W    9.973725e-01
## normtest.p    7.811050e-01
```

```
# residuals distribution: good. Raw scores/2.SE < 1

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_freq_CELEX_lem_concs)
```

```
## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##         1.009928         1.241384         1.249539
```

```
mean(vif(fit_freq_CELEX_lem_concs))
```

```
## [1] 1.16695
```

```r
1/vif(fit_freq_CELEX_lem_concs)
```

```
## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##        0.9901697        0.8055526        0.8002952
```

```r
# RESULTS: all good

step_freq_CELEX_lem_concs_AIC <- stepAIC(fit_freq_CELEX_lem_concs, direction="both")
```

```
## Start:  AIC=-7.96
## concs$s_freq_CELEX_lem ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                      Df Sum of Sq    RSS     AIC
## - concs$s_Haptic    1    0.7423 384.09 -9.1927
## <none>                          383.35 -7.9645
## - concs$s_Visual    1    4.5620 387.91 -5.2444
## - concs$s_Auditory  1    8.7073 392.06 -1.0032
##
## Step:  AIC=-9.19
## concs$s_freq_CELEX_lem ~ concs$s_Auditory + concs$s_Visual
##
##                      Df Sum of Sq    RSS     AIC
## <none>                          384.09 -9.1927
## + concs$s_Haptic    1    0.7423 383.35 -7.9645
## - concs$s_Visual    1    3.8276 387.92 -7.2362
## - concs$s_Auditory  1    9.0221 393.12 -1.9288
```

```r
step_freq_CELEX_lem_concs_F <- stepAIC(fit_freq_CELEX_lem_concs, direction="both",
test="F")
```

```
## Start:  AIC=-7.96
## concs$s_freq_CELEX_lem ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value    Pr(F)
## - concs$s_Haptic    1    0.7423 384.09 -9.1927  0.7649 0.382346
## <none>                          383.35 -7.9645
## - concs$s_Visual    1    4.5620 387.91 -5.2444  4.7006 0.030749 *
## - concs$s_Auditory  1    8.7073 392.06 -1.0032  8.9719 0.002914 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-9.19
## concs$s_freq_CELEX_lem ~ concs$s_Auditory + concs$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value    Pr(F)
## <none>                          384.09 -9.1927
## + concs$s_Haptic    1    0.7423 383.35 -7.9645  0.7649 0.382346
## - concs$s_Visual    1    3.8276 387.92 -7.2362  3.9462 0.047665 *
## - concs$s_Auditory  1    9.0221 393.12 -1.9288  9.3018 0.002443 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(fit_freq_CELEX_lem_concs)

##
## Call:
## lm(formula = concs$s_freq_CELEX_lem ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual, data = concs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.59645 -0.62930  0.01094  0.68222  2.74741
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.0005927  0.0493208   0.012  0.99042
## concs$s_Auditory  0.1484357  0.0495560   2.995  0.00291 **
## concs$s_Haptic   -0.0479850  0.0548676  -0.875  0.38235
## concs$s_Visual    0.1190287  0.0549004   2.168  0.03075 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9851 on 395 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.03681,    Adjusted R-squared:  0.02949
## F-statistic: 5.032 on 3 and 395 DF,  p-value: 0.001966

# length: RC1 lexicals
fit_RC1_lexicals_concs <- lm(concs$s_RC1_lexicals ~ concs$s_Auditory + concs$s_Haptic
+ concs$s_Visual, data = concs)
stat.desc(fit_RC1_lexicals_concs$residuals, norm = TRUE)

##                          x
## nbr.val       3.830000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.790833e+00
## max           4.935675e+00
## range         6.726507e+00
## sum           2.078199e-14
## median       -1.912252e-01
## mean          5.428045e-17
## SE.mean       4.997352e-02
## CI.mean.0.95  9.825761e-02
## var           9.564860e-01
## std.dev       9.780010e-01
## coef.var      1.801755e+16
## skewness      1.096509e+00
## skew.2SE      4.397413e+00
## kurtosis      1.924309e+00
```

```
## kurt.2SE      3.868479e+00
## normtest.W    9.373125e-01
## normtest.p    1.267186e-11

# residuals distribution: skewed and kurtosed. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(concs$s_RC1_lexicals)

##    vars   n mean sd median trimmed  mad  min max range skew kurtosis   se
## X1    1 383    0  1  -0.23    -0.1 0.83 -1.7 4.9   6.6 1.14     1.84 0.05

concs$log_s_RC1_lexicals_concs <- log(3 + concs$s_RC1_lexicals)

fit_RC1_lexicals_concs <- lm(concs$log_s_RC1_lexicals ~ concs$s_Auditory +
concs$s_Haptic + concs$s_Visual, data = concs)

# check residuals again
stat.desc(fit_RC1_lexicals_concs$residuals, norm = TRUE)

##                            x
## nbr.val       3.830000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -7.573554e-01
## max           1.034698e+00
## range         1.792054e+00
## sum          -6.081941e-15
## median       -1.684797e-02
## mean         -1.589181e-17
## SE.mean       1.574504e-02
## CI.mean.0.95  3.095779e-02
## var           9.494806e-02
## std.dev       3.081364e-01
## coef.var     -1.938964e+16
## skewness      1.960662e-01
## skew.2SE      7.862989e-01
## kurtosis     -1.815582e-01
## kurt.2SE     -3.649903e-01
## normtest.W    9.946692e-01
## normtest.p    2.064935e-01

# good

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_RC1_lexicals_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##         1.005934         1.244426         1.246755
```

```r
mean(vif(fit_RC1_lexicals_concs))

## [1] 1.165705

1/vif(fit_RC1_lexicals_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##        0.9941011        0.8035830        0.8020825

# RESULTS: all good

step_RC1_lexicals_concs_AIC <- stepAIC(fit_RC1_lexicals_concs, direction="both")

## Start:  AIC=-894.75
## concs$log_s_RC1_lexicals ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                      Df Sum of Sq    RSS     AIC
## - concs$s_Visual    1   0.05502 36.325 -896.17
## <none>                          36.270 -894.75
## - concs$s_Haptic    1   0.37136 36.642 -892.84
## - concs$s_Auditory  1   1.08098 37.351 -885.50
##
## Step:  AIC=-896.17
## concs$log_s_RC1_lexicals ~ concs$s_Auditory + concs$s_Haptic
##
##                      Df Sum of Sq    RSS     AIC
## <none>                          36.325 -896.17
## + concs$s_Visual    1   0.05502 36.270 -894.75
## - concs$s_Haptic    1   0.63244 36.958 -891.55
## - concs$s_Auditory  1   1.05179 37.377 -887.23

step_RC1_lexicals_concs_F <- stepAIC(fit_RC1_lexicals_concs, direction="both",
test="F")

## Start:  AIC=-894.75
## concs$log_s_RC1_lexicals ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value     Pr(F)
## - concs$s_Visual    1   0.05502 36.325 -896.17  0.5749 0.4487959
## <none>                          36.270 -894.75
## - concs$s_Haptic    1   0.37136 36.642 -892.84  3.8805 0.0495776 *
## - concs$s_Auditory  1   1.08098 37.351 -885.50 11.2955 0.0008559 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-896.17
## concs$log_s_RC1_lexicals ~ concs$s_Auditory + concs$s_Haptic
##
##                      Df Sum of Sq    RSS     AIC F Value     Pr(F)
```

```
## <none>                            36.325 -896.17
## + concs$s_Visual    1   0.05502 36.270 -894.75  0.5749 0.4487959
## - concs$s_Haptic    1   0.63244 36.958 -891.55  6.6159 0.0104860 *
## - concs$s_Auditory  1   1.05179 37.377 -887.23 11.0028 0.0009974 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_RC1_lexicals_concs)
```

```
##
## Call:
## lm(formula = concs$log_s_RC1_lexicals ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual, data = concs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.75736 -0.21704 -0.01685  0.21379  1.03470
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.04781    0.01582  66.251  < 2e-16 ***
## concs$s_Auditory  0.05347    0.01591   3.361 0.000856 ***
## concs$s_Haptic   -0.03450    0.01751  -1.970 0.049578 *
## concs$s_Visual   -0.01345    0.01774  -0.758 0.448796
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3094 on 379 degrees of freedom
##   (28 observations deleted due to missingness)
## Multiple R-squared:  0.04695,    Adjusted R-squared:  0.03941
## F-statistic: 6.224 on 3 and 379 DF,  p-value: 0.0003907
```

```
# distinctiveness: RC3 lexicals
fit_RC3_lexicals_concs <- lm(concs$s_RC3_lexicals ~ concs$s_Auditory + concs$s_Haptic
+ concs$s_Visual, data = concs)
stat.desc(fit_RC3_lexicals_concs$residuals, norm = TRUE)
```

```
##                          x
## nbr.val       3.830000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -1.701688e+00
## max           4.550983e+00
## range         6.252670e+00
## sum          -1.091835e-14
## median       -2.680258e-01
## mean         -2.845592e-17
## SE.mean       4.911158e-02
## CI.mean.0.95  9.656288e-02
## var           9.237760e-01
## std.dev       9.611326e-01
```

```
## coef.var     -3.377619e+16
## skewness      1.673295e+00
## skew.2SE      6.710544e+00
## kurtosis      3.357824e+00
## kurt.2SE      6.750305e+00
## normtest.W    8.551890e-01
## normtest.p    2.367318e-18
```

```
# residuals distribution: skewed and kurtosed. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(concs$s_RC3_lexicals)
```

```
##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 383    0  1  -0.27   -0.17 0.59 -1.44 4.99  6.43 1.83     3.67
##      se
## X1 0.05
```

```
concs$log_s_RC3_lexicals <- log(3 + concs$s_RC3_lexicals)

fit_RC3_lexicals_concs <- lm(concs$log_s_RC3_lexicals ~ concs$s_Auditory +
concs$s_Haptic + concs$s_Visual, data = concs)

# check residuals again
stat.desc(fit_RC3_lexicals_concs$residuals, norm = TRUE)
```

```
##                          x
## nbr.val       3.830000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -6.271740e-01
## max           9.026380e-01
## range         1.529812e+00
## sum          -4.562323e-15
## median       -5.616343e-02
## mean         -1.190429e-17
## SE.mean       1.395190e-02
## CI.mean.0.95  2.743214e-02
## var           7.455311e-02
## std.dev       2.730442e-01
## coef.var     -2.293663e+16
## skewness      8.766790e-01
## skew.2SE      3.515812e+00
## kurtosis      7.122003e-01
## kurt.2SE      1.431751e+00
## normtest.W    9.479679e-01
## normtest.p    2.347627e-10
```

```
# better though still non-normal

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
```

```r
# tolerance (pref. > 0.2)
vif(fit_RC3_lexicals_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##         1.005934         1.244426         1.246755

mean(vif(fit_RC3_lexicals_concs))

## [1] 1.165705

1/vif(fit_RC3_lexicals_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##        0.9941011        0.8035830        0.8020825

# RESULTS: all good

step_RC3_lexicals_concs_AIC <- stepAIC(fit_RC3_lexicals_concs, direction="both")

## Start:  AIC=-987.36
## concs$log_s_RC3_lexicals ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                    Df Sum of Sq    RSS     AIC
## - concs$s_Visual    1   0.02418 28.503 -989.04
## <none>                          28.479 -987.36
## - concs$s_Auditory  1   0.43934 28.919 -983.50
## - concs$s_Haptic    1   1.64314 30.122 -967.88
##
## Step:  AIC=-989.04
## concs$log_s_RC3_lexicals ~ concs$s_Auditory + concs$s_Haptic
##
##                    Df Sum of Sq    RSS     AIC
## <none>                          28.503 -989.04
## + concs$s_Visual    1   0.02418 28.479 -987.36
## - concs$s_Auditory  1   0.45652 28.960 -984.95
## - concs$s_Haptic    1   1.82968 30.333 -967.21

step_RC3_lexicals_concs_F <- stepAIC(fit_RC3_lexicals_concs, direction="both",
test="F")

## Start:  AIC=-987.36
## concs$log_s_RC3_lexicals ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                    Df Sum of Sq    RSS     AIC F Value     Pr(F)
## - concs$s_Visual    1   0.02418 28.503 -989.04  0.3218   0.57085
## <none>                          28.479 -987.36
## - concs$s_Auditory  1   0.43934 28.919 -983.50  5.8467   0.01608 *
## - concs$s_Haptic    1   1.64314 30.122 -967.88 21.8668 4.071e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Step:  AIC=-989.04
## concs$log_s_RC3_lexicals ~ concs$s_Auditory + concs$s_Haptic
##
##                     Df Sum of Sq    RSS     AIC F Value     Pr(F)
## <none>                           28.503 -989.04
## + concs$s_Visual    1   0.02418 28.479 -987.36  0.3218   0.57085
## - concs$s_Auditory  1   0.45652 28.960 -984.95  6.0862   0.01406 *
## - concs$s_Haptic    1   1.82968 30.333 -967.21 24.3927 1.179e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(fit_RC3_lexicals_concs)
```

```
##
## Call:
## lm(formula = concs$log_s_RC3_lexicals ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual, data = concs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62717 -0.18400 -0.05616  0.13150  0.90264
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.054305   0.014015  75.229  < 2e-16 ***
## concs$s_Auditory -0.034090   0.014098  -2.418   0.0161 *
## concs$s_Haptic    0.072572   0.015520   4.676 4.07e-06 ***
## concs$s_Visual   -0.008916   0.015716  -0.567   0.5708
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2741 on 379 degrees of freedom
##   (28 observations deleted due to missingness)
## Multiple R-squared:  0.07665,    Adjusted R-squared:  0.06934
## F-statistic: 10.49 on 3 and 379 DF,  p-value: 1.216e-06
```

```r
# freq: RC2 lexicals
fit_RC2_lexicals_concs <- lm(concs$s_RC2_lexicals ~ concs$s_Auditory + concs$s_Haptic
 + concs$s_Visual, data = concs)
stat.desc(fit_RC2_lexicals_concs$residuals, norm = TRUE)
```

```
##                          x
## nbr.val       3.830000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -2.701781e+00
## max           2.799024e+00
## range         5.500804e+00
## sum          -9.728329e-15
## median       -5.204421e-03
```

```
## mean         -2.532129e-17
## SE.mean       4.939207e-02
## CI.mean.0.95  9.711437e-02
## var           9.343578e-01
## std.dev       9.666218e-01
## coef.var     -3.817428e+16
## skewness      1.177816e-01
## skew.2SE      4.723486e-01
## kurtosis     -3.253570e-02
## kurt.2SE     -6.540720e-02
## normtest.W    9.975488e-01
## normtest.p    8.476297e-01

# residuals distribution: good. Raw scores/2.SE < 1

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_RC2_lexicals_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##         1.005934         1.244426         1.246755

mean(vif(fit_RC2_lexicals_concs))

## [1] 1.165705

1/vif(fit_RC2_lexicals_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##        0.9941011        0.8035830        0.8020825

# RESULTS: all good

step_RC2_lexicals_concs_AIC <- stepAIC(fit_RC2_lexicals_concs, direction="both")

## Start:  AIC=-19.01
## concs$s_RC2_lexicals ~ concs$s_Auditory + concs$s_Haptic + concs$s_Visual
##
##                    Df Sum of Sq    RSS      AIC
## - concs$s_Haptic    1    0.5196 357.44 -20.4482
## <none>                          356.92 -19.0054
## - concs$s_Visual    1    3.1284 360.05 -17.6631
## - concs$s_Auditory  1   20.6663 377.59   0.5524
##
## Step:  AIC=-20.45
## concs$s_RC2_lexicals ~ concs$s_Auditory + concs$s_Visual
##
##                    Df Sum of Sq    RSS      AIC
## <none>                          357.44 -20.4482
## - concs$s_Visual    1    2.6135 360.06 -19.6580
```

```
## + concs$s_Haptic    1    0.5196 356.92 -19.0054
## - concs$s_Auditory  1   21.1095 378.55  -0.4722
```

```
step_RC2_lexicals_concs_F <- stepAIC(fit_RC2_lexicals_concs, direction="both",
test="F")
```

```
## Start:  AIC=-19.01
## concs$s_RC2_lexicals ~ concs$s_Auditory + concs$s_Haptic + concs$s_Visual
##
##                    Df Sum of Sq    RSS      AIC F Value     Pr(F)
## - concs$s_Haptic    1    0.5196 357.44 -20.4482  0.5518   0.45806
## <none>                          356.92 -19.0054
## - concs$s_Visual    1    3.1284 360.05 -17.6631  3.3219   0.06915 .
## - concs$s_Auditory  1   20.6663 377.59   0.5524 21.9445 3.918e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=-20.45
## concs$s_RC2_lexicals ~ concs$s_Auditory + concs$s_Visual
##
##                    Df Sum of Sq    RSS      AIC F Value     Pr(F)
## <none>                          357.44 -20.4482
## - concs$s_Visual    1    2.6135 360.06 -19.6580  2.7785   0.09636 .
## + concs$s_Haptic    1    0.5196 356.92 -19.0054  0.5518   0.45806
## - concs$s_Auditory  1   21.1095 378.55  -0.4722 22.4415 3.065e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_RC2_lexicals_concs)
```

```
##
## Call:
## lm(formula = concs$s_RC2_lexicals ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual, data = concs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7018 -0.6471 -0.0052  0.6136  2.7990
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.006071   0.049614  -0.122   0.9027
## concs$s_Auditory  0.233808   0.049911   4.684 3.92e-06 ***
## concs$s_Haptic   -0.040811   0.054942  -0.743   0.4581
## concs$s_Visual    0.101407   0.055638   1.823   0.0691 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9704 on 379 degrees of freedom
##   (28 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.06564,    Adjusted R-squared:  0.05825
## F-statistic: 8.875 on 3 and 379 DF,  p-value: 1.067e-05

# additional var: age of acquisition
fit_AoA_Brysbaertetal2014_concs <- lm(concs$s_AoA_Brysbaertetal2014 ~
concs$s_Auditory + concs$s_Haptic + concs$s_Visual, data = concs)
stat.desc(fit_AoA_Brysbaertetal2014_concs$residuals, norm = TRUE)

##                          x
## nbr.val       3.980000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -3.025869e+00
## max           2.147610e+00
## range         5.173479e+00
## sum           1.313359e-14
## median       -1.358284e-02
## mean          3.302938e-17
## SE.mean       4.390837e-02
## CI.mean.0.95  8.632199e-02
## var           7.673221e-01
## std.dev       8.759693e-01
## coef.var      2.652091e+16
## skewness     -8.877953e-02
## skew.2SE     -3.628914e-01
## kurtosis     -2.872845e-01
## kurt.2SE     -5.885934e-01
## normtest.W    9.957891e-01
## normtest.p    3.666147e-01

# residuals distribution: good. Raw scores/2.SE < 1

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_AoA_Brysbaertetal2014_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##         1.006555         1.256294         1.261234

mean(vif(fit_AoA_Brysbaertetal2014_concs))

## [1] 1.174694

1/vif(fit_AoA_Brysbaertetal2014_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##        0.9934875        0.7959923        0.7928746

# RESULTS: all good

step_AoA_Brysbaertetal2014_concs_AIC <- stepAIC(fit_AoA_Brysbaertetal2014_concs,
direction="both")
```

```
## Start:  AIC=-98.41
## concs$s_AoA_Brysbaertetal2014 ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                      Df Sum of Sq    RSS     AIC
## <none>                            304.63 -98.411
## - concs$s_Auditory  1    4.3063 308.93 -94.824
## - concs$s_Haptic    1   20.6142 325.24 -74.350
## - concs$s_Visual    1   27.0765 331.70 -66.520
```

```
step_AoA_Brysbaertetal2014_concs_F <- stepAIC(fit_AoA_Brysbaertetal2014_concs,
direction="both", test="F")
```

```
## Start:  AIC=-98.41
## concs$s_AoA_Brysbaertetal2014 ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                      Df Sum of Sq    RSS     AIC F Value     Pr(F)
## <none>                            304.63 -98.411
## - concs$s_Auditory  1    4.3063 308.93 -94.824   5.570   0.01876 *
## - concs$s_Haptic    1   20.6142 325.24 -74.350  26.662 3.854e-07 ***
## - concs$s_Visual    1   27.0765 331.70 -66.520  35.020 7.097e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_AoA_Brysbaertetal2014_concs)
```

```
##
## Call:
## lm(formula = concs$s_AoA_Brysbaertetal2014 ~ concs$s_Auditory +
##     concs$s_Haptic + concs$s_Visual, data = concs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.02587 -0.58392 -0.01358  0.66456  2.14761
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.007778   0.044092   0.176   0.8601
## concs$s_Auditory -0.104344   0.044213  -2.360   0.0188 *
## concs$s_Haptic   -0.254806   0.049347  -5.164 3.85e-07 ***
## concs$s_Visual   -0.294074   0.049693  -5.918 7.10e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8793 on 394 degrees of freedom
##   (13 observations deleted due to missingness)
## Multiple R-squared:  0.2327, Adjusted R-squared:  0.2268
## F-statistic: 39.82 on 3 and 394 DF,  p-value: < 2.2e-16
```

```r
# additional var: concreteness
fit_concrete_Brysbaertetal2014_concs <- lm(concs$s_concrete_Brysbaertetal2014 ~
concs$s_Auditory + concs$s_Haptic + concs$s_Visual, data = concs)
stat.desc(fit_concrete_Brysbaertetal2014_concs$residuals, norm = TRUE)

##                           x
## nbr.val       3.980000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -2.047456e+00
## max           2.338430e+00
## range         4.385886e+00
## sum           1.426810e-14
## median       -3.294205e-02
## mean          3.585668e-17
## SE.mean       3.940998e-02
## CI.mean.0.95  7.747834e-02
## var           6.181523e-01
## std.dev       7.862266e-01
## coef.var      2.192692e+16
## skewness      2.636678e-01
## skew.2SE      1.077757e+00
## kurtosis     -8.873867e-02
## kurt.2SE     -1.818093e-01
## normtest.W    9.907920e-01
## normtest.p    1.382913e-02

# residuals distribution: skew. Raw scores/2.SE > 1
# have to log-transform DV and re-run regression

psych::describe(concs$s_concrete_Brysbaertetal2014)

##    vars   n mean sd median trimmed  mad   min  max range skew kurtosis
## X1    1 398    0  1  -0.21   -0.04 1.11 -1.71 1.85  3.56 0.38    -1.13
##      se
## X1 0.05

concs$log_s_concrete_Brysbaertetal2014 <- log(3 + concs$s_concrete_Brysbaertetal2014)

fit_concrete_Brysbaertetal2014_concs <- lm(concs$log_s_concrete_Brysbaertetal2014 ~
concs$s_Auditory + concs$s_Haptic + concs$s_Visual, data = concs)

# check residuals again
stat.desc(fit_concrete_Brysbaertetal2014_concs$residuals, norm = TRUE)

##                           x
## nbr.val       3.980000e+02
## nbr.null      0.000000e+00
## nbr.na        0.000000e+00
## min          -7.098116e-01
## max           8.091234e-01
```

```
## range         1.518935e+00
## sum          -2.052178e-15
## median        2.483684e-02
## mean         -5.137822e-18
## SE.mean       1.358202e-02
## CI.mean.0.95  2.670168e-02
## var           7.341959e-02
## std.dev       2.709605e-01
## coef.var     -5.273840e+16
## skewness     -8.997464e-03
## skew.2SE     -3.677765e-02
## kurtosis     -2.456084e-01
## kurt.2SE     -5.032068e-01
## normtest.W    9.933368e-01
## normtest.p    7.598647e-02

# good

# Check multicollinearity: largest VIF (pref. < 10), mean VIF (pref. around 1), and
# tolerance (pref. > 0.2)
vif(fit_concrete_Brysbaertetal2014_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##         1.006555         1.256294         1.261234

mean(vif(fit_concrete_Brysbaertetal2014_concs))

## [1] 1.174694

1/vif(fit_concrete_Brysbaertetal2014_concs)

## concs$s_Auditory   concs$s_Haptic   concs$s_Visual
##        0.9934875        0.7959923        0.7928746

# RESULTS: all good

step_concrete_Brysbaertetal2014_concs_AIC <-
stepAIC(fit_concrete_Brysbaertetal2014_concs, direction="both")

## Start:  AIC=-1032.4
## concs$log_s_concrete_Brysbaertetal2014 ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                     Df Sum of Sq    RSS      AIC
## <none>                           29.148 -1032.40
## - concs$s_Auditory  1    1.1547 30.302 -1018.94
## - concs$s_Haptic    1    3.5383 32.686  -988.80
## - concs$s_Visual    1    4.8804 34.028  -972.79

step_concrete_Brysbaertetal2014_concs_F <-
stepAIC(fit_concrete_Brysbaertetal2014_concs, direction="both", test="F")
```

```
## Start:  AIC=-1032.4
## concs$log_s_concrete_Brysbaertetal2014 ~ concs$s_Auditory + concs$s_Haptic +
##     concs$s_Visual
##
##                    Df Sum of Sq    RSS      AIC F Value     Pr(F)
## <none>                          29.148 -1032.40
## - concs$s_Auditory  1    1.1547 30.302 -1018.94  15.609 9.230e-05 ***
## - concs$s_Haptic    1    3.5383 32.686  -988.80  47.828 1.888e-11 ***
## - concs$s_Visual    1    4.8804 34.028  -972.79  65.971 5.946e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit_concrete_Brysbaertetal2014_concs)
```

```
##
## Call:
## lm(formula = concs$log_s_concrete_Brysbaertetal2014 ~ concs$s_Auditory +
##     concs$s_Haptic + concs$s_Visual, data = concs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.70981 -0.21810  0.02484  0.19189  0.80912
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.04056    0.01364  76.294  < 2e-16 ***
## concs$s_Auditory -0.05403    0.01368  -3.951 9.23e-05 ***
## concs$s_Haptic    0.10557    0.01526   6.916 1.89e-11 ***
## concs$s_Visual    0.12485    0.01537   8.122 5.95e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.272 on 394 degrees of freedom
##   (13 observations deleted due to missingness)
## Multiple R-squared:  0.3571, Adjusted R-squared:  0.3522
## F-statistic: 72.94 on 3 and 394 DF,  p-value: < 2.2e-16
```

```
# Results: Iconicity of concepts and comparison with properties:
# The properties sample was characterized by smaller advantages for Auditory
# predictor, compared to the concepts sample. The tendency of either larger or
# opposite scores for the Auditory strength was less evident, even though it was
# still marginally present. This raw-figure difference was not statistically tested.


# END
```