Applied Corpus Linguistics 2 (2022) 100022

Contents lists available at ScienceDirect

Applied Corpus Linguistics

journal homepage: www.elsevier.com/locate/acorp

A tutorial on norming linguistic stimuli for clinical populations

Oliver Delgaram-Nejad[a,*], Gerasimos Chatzidamianos[a], Dawn Archer[a], Alex Bartha[b], Louise Robinson[c]

[a] Manchester Metropolitan University, Manchester, United Kingdom
[b] North-East London NHS Foundation Trust, Greater London, United Kingdom
[c] Lancashire and South Cumbria NHS Foundation Trust, Lancashire, University of Manchester, Manchester, United Kingdom

A B S T R A C T

Stimuli norming (the process of controlling experimental items to minimise bias) is important for the validity of psycholinguistic experiments. Survey norming (asking large numbers of people to rate or otherwise define the items) is typically used for this purpose but requires large samples. Clinical populations are not always large, nor easy to reach. Clinical participants often have ongoing symptomatology, and some cohorts experience language and communication difficulties. We present a corpus-linguistic method suitable for clinical populations for which survey norming is difficult or inappropriate. We also include the experiment generated, which measures metaphor-creation behaviour in schizophrenia to test Cognitive Constraint Theory (CCT) in clinical and nonclinical populations (see S2.1). We describe the design rationale before outlining the design stages in tutorial form. This allows us to show readers why the approach was needed and support them to consider and respond to the challenges that we encountered. We conclude that it is easier to consider norming and design practices in parallel when experimental units are defined linguistically. Corpus stimuli norming provides a versatile alternative when survey norming is prohibitive, especially in speech pathology.

1. Introduction

This paper is both a case for greater corpus norming (the name we give to the process for which we're arguing) uptake and an illustrative tutorial for those seeking to involve applied linguists within their experimental work. Our primary audience is applied (corpus) linguists, whose expertise is of particular value to experimental psycholinguists. We ask readers to view this as a position paper first and tutorial second, as our main argument is about the value of linguistic corpora in the design of language experiments. This generally calls for more dialogue between applied corpus linguists and experimental psychologists and psycholinguists, especially when engaging clinical questions and populations. The tutorial component is our way of providing evidence of a working result that also confers practical benefits to the wider research community. Note that the particulars of this experimental design relate to hypotheses of whether people with a diagnosis of schizophrenia show preferences in the syntactic variation of literary devices and whether these are consistent with those shown to some extent in nonclinical populations.

Most language use occurs in the field. Most language measurement occurs in the lab. Language loses its everyday context in the lab (Clark and Bangerter, 2004), and so language experiments suffer problems of ecological validity. Clinical language studies are primarily interested in how pathology affects everyday language, rather than task performance in controlled contexts. Our aim was a controlled experiment suitable for participants with a diagnosis of schizophrenia that would advance our understanding of everyday language use in this cohort. We are particularly interested in whether metaphor comprehension influences formal thought disorder or incoherent speech (McKenna and Oh, 2005; Demjaha et al. 2017). We chose to begin stimuli selection with items present in linguistic corpora, so that items were taken from an everyday usage context. This approach is similar to survey norming, but the order of stimuli creation is reversed. In survey norming, stimuli are designed and then presented to large samples who rate various attributes, such as subjective visual complexity for picture items. Our corpus-norming approach filters language data derived from large samples to create stimuli. It is ideal for situations where time, cost, and feasibility limitations prohibit a standalone norming study. Corpus norming is no less reproducible, because searches (keywords, dates, full results) are easily reported and compared. See Burnard's (2002) appraisal of British National Corpus (BNC; 2007) construction for other metrics.

Reproducibility is especially important for experiments based on Relevance Theory, where pragmatically different stimuli generate mixed findings (Nicolle &Clarke, 1999; Van Der Henst and Sperber, 2004). This is important because such experiments are used to study pragmatic disability. Corpus norming also provides insights about the patterns that language users tend to overlook, providing data that survey approaches cannot. Our main suggestion is that corpus norming shows potential and would benefit from greater uptake.

Our reasons for writing this tutorial are closely tied to our reasons for developing the method. The problem is a lack of *interdisciplinary* design. We required an experiment that combined linguistic, psychological/psycholinguistic, and psychiatric viewpoints. We developed our own because we could not find an extant example relevant to our research question. This may have to do with language experiments being

* Corresponding author at: Department of Linguistics, Manchester Metropolitan University, Manchester M15 6LL, United Kingdom.
E-mail address: o.delgaram-nejad@mmu.ac.uk (O. Delgaram-Nejad).

https://doi.org/10.1016/j.acorp.2022.100022
Received 18 April 2022; Received in revised form 14 June 2022; Accepted 28 June 2022
Available online 2 July 2022

uncommon in linguistics (Gilquin and Gries, 2009). Linguistic theory is also uncommon in experimental psychology. Design errors in language experiments tend to be mainly linguistic or methodological for these reasons. Linguists and psychologists are trying to do independently what requires the sum of their knowledge and skills. This manifests conceptually as one reference to 'experimentation' in *The Oxford Handbook of Cognitive Linguistics* (Geeraerts & Cuyckens, 2007) and no reference to 'corpora' in the *Handbook of Psycholinguistics* (Fernández et al., 2018). It occurs practically in examples like Ferber (1995), where reliability and validity issues were identified in a corpus of verbal slips. These could have been anticipated during construction with psychological input.

A handful of studies have successfully used corpora to support experimental studies directly or indirectly. Much of this small body of work has been about the latter. Schreuder and Kerkman (1987) highlighted the importance of word frequency for language cognition and corpora as reliable sources of that information. The London-Lund Corpus was used to generalise estimates of speech error rates overall (Garnham et al., 1982) and production mechanisms for adults and children (Wijnen, 1992) and aphasic and non-aphasic adults (Schwartz et al., 1994).

More recent and direct applications come from experimental semantics and pragmatics (Bezuidenhout, 2010; Archer and Grundy, 2011). Many of these are experimental designs motivated by corpus findings. This makes sense because corpus analyses generate correlational findings and causal findings require experimentation (De Ruiter, 2013). Spenader's (2002) corpus study of presupposition accommodation, where 'my dog…' is accommodated by presupposing *has a dog*, recommended further work on presupposition triggers: *too*, the aspectual adverbs *still* and *already*, and definite noun phrases. This motivated experimental work on reading times and acceptability ratings. Examples include acceptability ratings of the returned action triggers *the* and *too* by context plausibility (Singh et al. 2015), acceptability ratings for several trigger types by presuppositional context, (Tiemann et al., 2011), and reading time by varied noun phrase definiteness and context plausibility (L. Frazier, 2006). This extends to syntax studies of antecedent selection. Corpus work (Arnold, 1998) on preferential resolution for first-position or subject referents (such as *he* denoting *Mark* in *Mark called Dave. He was upset.*) motivated clinical (autism) narrative elicitation (Arnold, 2009) and general pronoun resolution and sentence continuation tasks (Kehler, 2008). Related corpus work checked whether an antecedent's properties influence accented pronouns, before a production experiment on whether they signal topic shifts (Wolters and Beaver, 2001). Related production experiments include sentence continuation tasks (Stevenson et al., 1994), manipulations of speaker attention (Brennan, 1995), and oral story continuation combined with post-hoc corpus analyses (Arnold, 2001).

Corpus data have also challenged causational studies. Data on infants' mental-state verb use, namely differences at ages two and three between conversational and genuine references to mental states, challenged the design of false-belief tasks (Shatz et al.1983; Bartsch and Wellman, 1995). A comparison of three conversational corpora, favouring signal response over anticipation models of turn-taking, challenged experimental findings on floor transfer onsets (FTOs), or timings calculated by subtracting offset from onset times (Heldner and Edlund, 2010). Further, corpus studies associate turn endings with completion points: syntactic and intonational (Caspers, 2003), complete in grammar or sound, and pragmatic, complete in purpose (Ford & Thompson, 1996). Corpus findings raise questions for experimental testing and replication issues benefit from corpus investigation.

Other uses include hypothesis formation (Westera and Brasoveanu, 2014), developing a typology (Sassoon, 2012) and acceptability ratings (Sassoon, 2013) for theory-building, ruling out explanations for experimental findings (Van Tiel et al. 2016), comparing child and adult speech (Eiteljoerge et al., 2018), and developing natural language stimuli (Degen, 2015). Some work has looked at extrapolating psycholinguistic variables from corpora using machine learning approaches (Mandera et al., 2015). This is however very different from what we describe in Section 2.

The above approaches are rarer in health research. Corpus health applications include metaphor use in chronic pain (Munday et al., 2020), metaphoricity in schizophrenia (Bilgrami et al., 2020), comparing the Linguistic Inquiry and Word Count (LIWC; Pennebaker, 2017) internal dictionary to human pain ratings (Ziemer and Korkmaz, 2017), and pronoun use in patient information sheets (Isaacs et al., 2020). There is much less on developing natural language stimuli for clinical cohorts.

Researchers have been enhancing experiments with corpora for some time. The attached literature is however limited. There is even less work on the processes required to situate corpus-derived stimuli within broader study designs. This is before considering the requirements of clinical populations. The limited literature is nonetheless diverse and demonstrates how versatile corpora can be. We hope that our tutorial description and materials encourage discussion and further developments.

2. Tutorial

2.1. Overview

We cover our design in three stages: stimuli selection, building stimuli lists, and fitting lists to the experimental design. Our experience was not this linear. We worked across stages and encourage readers to do the same, although it is helpful to partition things in this way where possible. We also stress the value of record keeping.

Our experiment is about sensory metaphors and whether language pathology has any effect on them. We specifically looked at synaesthetic metaphors. Synaesthesia describes both a neurological condition and a literary technique. Linguistic synaesthesia occurs when words relating to discrete senses are grammatically combined. The example *smelly light* is a synaesthesia because the olfactory adjective modifies the visual noun. Metaphors are created by mapping source qualities onto targets.[1]

We take a Conceptual Mapping Theory (CMT) approach to metaphor (Lakoff and Johnson, 1980; Gibbs and Ortony, 2008) over analogical or categorisation approaches (Holyoak and Stamenković, 2018) because our experiment tests CCT (Shen, 2002). The theory suggests that people prefer synaesthesias where a less salient (as the term is used in CMT, cf. corpus linguisitcs) sense modifies a more salient one. CCT argues that source-target mappings adhere to a sensory hierarchy in which vision is most salient, followed by sound, smell, taste, and touch. An example of such a preference would be using *warm/touch* to modify *glow/vision* (rather than the reverse). CCT suggests a *cognitive* rather than contextual basis because corpus work locates these preferences across languages, historical/temporal, and stylistic contexts. It however lacks sufficient experimental support. We therefore designed our experiment to permit a test of CCT in the nonclinical sample (to explore CCT's claims) as well as a broader group comparison of syntactic preferences for linguistic synaesthesia in clinical and nonclinical groups independent of CCT (to explore FTD in schizophrenia).

The example *smelly opinion* is a metaphor because *opinion* borrows the olfactory qualities of *smelly*. A synaesthetic metaphor is a construc-

[1] We appreciate that some definitions of metaphor require not only a source-target mapping but also for source characteristics to meaningfully contribute to understanding the target. *Smelly opinion* is therefore only a metaphor by these definitions if the qualities of *smelly* inform an understanding of *opinion*. We operationalised sensory metaphors on source-target relations alone because our experiment assesses (1) how corpus norming informs past research on syntactic preferences under a conceptual mapping approach and (2) whether individuals with a diagnosis of schizophrenia display preferences comparable to individuals with no known psychiatric history. It also made sense to hold this dimension constant because schizophrenia is associated with semanto-pragmatic difficulties that are currently difficult to control due to their heterogeneity. Our experiment could be used to explore participant approaches to *ground* across prominent theories of metaphor.

tion that adheres to the rules of both. We operationalised synaesthetic metaphors by splitting them into primes and responses. This allowed us to blind participants to their metaphor-creation. This is what our experiment does at the theoretical level. At the practical level, we measured preferences for combining synaesthetic pairs of sensory adjectives with concrete or abstract nouns in people with and without a diagnosis of schizophrenia. Most of our design work went into stimuli selection because sensory adjectives are sensitive to many psycholinguistic confounders (Huisman and Majid, 2018). We also considered that schizophrenia cohorts are small and hard to recruit, not least because so few people meet clinical (diagnosis of schizophrenia) *and* experimental (well enough to take part) criteria. This is why we talk about stimuli selection first and why it forms the bulk of our approach.

2.2. Selecting stimuli

We select control variables whenever we select stimuli. Researchers are known to select the former intuitively (Noveck and Sperber, 2004). We discourage this because it biases data. Some of the most popular picture-naming controls affect naming speed (Perret and Bonin, 2019) for example. We encourage readers to instead predict confounders and work backwards systematically, consolidating as much as possible. Our priority was accessible vocabulary. We wanted to avoid results based on lack of understanding. Vocabulary is important in schizophrenia and other conditions that affect semantic memory (Hwang et al., 2021). We operationalised accessibility through word frequency, because words common in everyday speech *tend to be* acquired earlier (Pagel and Meade, 2018). We avoid saying that we operationalised accessible vocabulary as age of acquisition because word frequency and age of acquisition exert different effects (Dewhurst et al., 1998). High-frequency words also benefit from high-readability and processing fluency (Chen and Meurers, 2018). We began with potential sources of *linguistic* bias.

We then focused on our perceptual adjectives. These are difficult to operationalise. Non-visual types (*sound, smell, taste, touch*) lack construct validity. This shows in picture tasks: representing *tall* is easier than *salty*. The same applies to internal states (*happy*) and aesthetic evaluations (*beautiful*). Context-of-use is also important. *Beautiful* can relate to visual or auditory aesthetics. *Bitter* can describe a taste or internal state. These challenges prompt inventive responses, like representing *beautiful* in terms of modified facial symmetry (Liao et al., 2016; Liao and Meskin, 2017).

We wanted maximum accessibility, construct validity, and representation ease with minimal polysemy. Our approach resembled a waterfall process (Bomarius et al., 2009). We essentially listed potential limitations, ranked them, and used this to determine our item-development priorities. Accessible vocabulary was the highest priority, and so we began there. We first searched for accessible *sensory* adjectives in the vocabulary level lists (Nation, 2004) provided with AntWordProfiler 1.5.1. (Antony, 2021). These lists are derived from the most common word families in the British National Corpus (BNC, 2007). We found a few in the higher vocabulary brackets. There were not enough, and they were quite inaccessible. This step also revealed the need for an operational definition for *sensory* adjectives. We then stepped back from vocabulary profiling based on 10,000 word families and instead looked at word frequency across the BNC at large. The BNC allows users to generate custom frequency lists. We saw this as an opportunity to introduce experimental controls. BNC adjectives are tagged as general (AJ0), comparative (AJC), and superlative (AJS). We excluded AJC types because comparatives draw attention to their counterparts. We excluded AJS types because scalar endpoints emphasise one end of a scale. This gave us a list of the most frequent *general* adjectives in the BNC. The BNC categorises nouns as number neutral common (NN0), singular common (NN1), plural common (NN2), and singular and plural proper (NP0). We excluded all but *singular common* to rule out variations in grammatical number and references to proper single entities. This gave us a list of the most frequent *singular common* nouns in the BNC.

Replicating this step:

- register to use BNCweb (http://bncweb.lancs.ac.uk/bncwebSignup/user/login.php)
- login to BNCweb
- click ‘frequency lists’
- choose ‘AJ0’ from POS tag list
- select ‘containing’ for ‘word pattern’
- select ‘Whole BNC’ for ‘range of texts’
- click ‘show list’
- select ‘Download whole Frequency List’ from the ‘New Frequency List’ dropdown
- click ‘go’
- repeat the above steps for ‘NN1’

We then extended this approach to address other limitations. This is what we mean by consolidation. Our next stage was about defining *sensory* adjectives. We used the UCREL Semantic Analysis System (USAS; Rayson et al., 2004), within Wmatrix 4 (Rayson, 2008). USAS is a tool for automatic semantic analysis. Its taxonomy of 21 major discourse fields and 232 categories contains *taste* (X3.1), *sound* (X3.2), *touch* (X3.3), *sight* (X3.4), and *smell* (X3.5) subcategories. It assigns these tags using robust methods (see Rayson et al., 2004). We screened our adjective frequency list for the sensory tags above. This gave us a reduced frequency list of *sensory* adjectives. We also screened our *singular common* noun frequency list for *general object* (O2) and *thought, belief* (X2.1) subcategories. This gave us a reduced frequency list of *semantically concrete* and *abstract* nouns. We chose single subcategories rather than groups of related subcategories to minimise noise.

Replicating this step[2]:

- open the downloaded ‘AJ0’ and ‘NN1’ frequency lists
- copy list content into spreadsheet software
- extract word column only (without ‘word’ header)
- create new word-only lists
- register to use Wmatrix 4 or 5 (https://ucrel.lancs.ac.uk/wmatrix/)
- login to Wmatrix 4 or 5
- click ‘tag wizard’
- click ‘browse’ and select word-only ‘AJ0’ list
- click ‘upload now’
- download USAS tagged output (not CLAWS)
- extract only ‘AJ0’ list words tagged ‘X3.1′–‘X3.5′
- repeat the above steps for ‘NN1’
- extract only NN1 words tagged O2 and X2.1

We extended the approach again using LIWC (Pennebaker, 2017). LIWC is a sentiment analysis tool that uses an internal dictionary with psychometric properties (Pennebaker et al., 2015). Its taxonomy covers 82 variables including *affect* and *perceptual processes*. We excluded everything tagged with *affect* to rule out valence. We kept *sensory* adjectives and *general object* nouns tagged with *perceptual processes*. We excluded *thought, belief* nouns tagged with *perceptual processes*. This step yielded two further reduced frequency lists, controlled for valence and psychometric overlap.

Replicating this step[3]:

- register to use LIWC2015 or LIWC-22 (https://www.liwc.app)
- install/run LIWC2015 or LIWC-22
- click ‘category options’

[2] Note that this step was carried out in Wmatrix4 and is reported here as it was carried out at the time. Wmatrix5 has since been released and may involve slightly different steps.

[3] Note that this step was carried out in LIWC2015 and is reported here as it was carried out at the time. LIWC2022 has since been released and may involve slightly different steps.

- click 'select none'
- select 'affect' and 'perceptual processes'
- click 'OK'
- click 'categorise words'
- select word-only 'AJ0' list
- click 'file'
- choose 'save results'
- repeat the above steps for 'NN1'
- exclude and retain items accordingly

This procedure created two lists of potential stimuli with the following properties:

- highest BNC frequency *semantically* sensory, non-affective, *psychometrically* perceptual general adjectives
- highest BNC frequency *semantically* concrete, non-affective, *psychometrically* perceptual *general object* nouns
- highest BNC frequency *semantically* abstract, non-affective, *thought, belief* nouns

We draw reader attention to how each stage informed surrounding stages. Sensory was operationalised *semantically (USAS, sensory)* and *psychometrically* (LIWC, *perceptual processes*). Most important is how the LIWC stage (combined with others) allowed us to introduce a linguistic-pragmatic aspect between *sensory adjectives* and *general object* nouns, with *perceptual processes* operationalised as a source of shared contextual meaning. We operationalised contextual meaning *psychometrically* because our total sample included people with and without impaired semantic knowledge. Concrete was operationalised *semantically* (USAS, *general object*), *psychometrically* (LIWC, *perceptual processes*), and *pragmatically* (LIWC, *perceptual processes*). *Abstract* was operationalised *semantically* (USAS, *thought, belief*), *psychometrically* (LIWC, **not** *perceptual processes*), and *pragmatically* (LIWC, **not** *perceptual processes*).

We then revisited our research aims to decide how many adjectives and nouns to extract from these final lists.

2.3. Building stimuli lists

Our experiment measures blinded metaphor-creation preferences in clinical and non-clinical groups, and most of Section 2.2. was about that. We also wanted it to test CCT. CCT experiments have looked at synaesthetic metaphors both as word pairs and as within sentence contexts (Shen and Aisenman, 2008).

We broke the metaphor-creation task into prime and response components. We operationalised synaesthetic word pairs within the prime component. We paired *sensory* adjectives from different modalities (e.g. *touch* and *vision)* to form primes. We operationalised metaphoricity within the response component. Participant responses were about noun selection. This means that a prime functions as a synaesthetic word pair before the participant chooses a noun. It also means that the prime functions as a joint-modifier to the noun after the choice is made. The prime is non-metaphoric but synaesthetic in isolation. The nouns are non-metaphoric and non-synaesthetic in isolation. Choosing a noun transforms the prime and noun into synaesthetic metaphor components. Task engagement alters the stimuli properties for prime and response. CCT experiments suggest that the preference holds for both synaesthetic metaphors and (non-metaphoric) synaesthetic word pairs in sentence contexts (Shen and Aisenman, 2008). We therefore considered it reasonable to combine them into a single task. We needed to be able to vary syntactic/grammatical order easily because CCT is ultimately a syntactic theory. We found it easier to vary the order of the adjectives/primes and responses/nouns separately by using this metaphor-splitting approach. This brought us back to our final frequency lists and stimuli list construction.

We used tables to keep track and encourage readers to consider the same. This stage became increasingly complex and quick visual overviews resolved many points of confusion. We calculated permutations of the *sensory* subcategories. Permutations are for when order matters. This is true for syntax. We calculated permutations without repetition, meaning no duplicates (*smell-smell*). We had five *sensory* subcategories (n = 5) and wanted them in pairs (r = 2). This gave us 20 permutations. We separated them according to CCT. 10 followed the *preferred syntax* (*touch-sight*) and 10 the other (*sight-touch*). We spent considerable time working out how many adjectives were needed to ensure that each adjective (*visual*), adjective-pairing/prime (*visual quiet*), *sensory* subcategory (*sight*), and *sensory* subcategory pairing (*sight-sound*) repeated equally, wanting to avoid exposure effects. Our tables were especially useful here. We extracted the three highest-frequency adjectives for each *sensory* subcategory from our *adjective* frequency list (see Tables 1–3 in Appendix I). Some top-three tokens shared a lemma, like *sweet* and *sweetened*. We chose one token per lemma when this happened, namely the higher frequency one. We chose only three adjectives because (1) it was the minimum required for equal exposures given our permutations and (2) some *sensory* subcategories had only three adjectives left after Section 2.2. This gave us a list of *sensory* adjectives (Table 1), a list of 90 primes in the *preferred syntax* (Table 2), and a list of 90 primes in the *nonpreferred-syntax* (Table 3). Each *sensory* adjective repeats 12 times in Table 2 and 12 times in Table 3, 24 overall. No *sensory* adjective pairing repeats. Each *sensory* subcategory repeats four times in Table 2 and four times in Table 3, eight overall. No *sensory* subcategory pairing repeats.

With adjectives, we calculated permutations of the *sensory* subcategories. We then calculated the fewest adjectives required for equal exposures. We then distributed the adjectives equally across the permutations. We calculated permutations for the nouns differently because we do not vary them across presentations in the same way as the adjectives. Participants always have a choice of concrete or abstract nouns, whereas they may see a *smell-taste* prime in one presentation and a *sight-touch* prime in the next. We repeated our adjective extraction method with the final frequency list for nouns, selecting three *general object* and three *thought, belief*. This gave us a list of six nouns (Table 4). We then realised the need to vary the noun syntax to avoid reading order and primacy effects. We calculated permutations for the nouns themselves, without repetition. We had six nouns (n = 6) and wanted them in pairs (r = 2). This gave us 30 permutations. 12 of these were single-subcategory (*edge boomerang* being *concrete-concrete*). We set those aside because the experiment requires one concrete and one abstract noun per presentation, not two of the same subcategory. This gave us 18 permutations. We separated them into two syntactic orders. Nine were *concrete first* (Table 5) and nine were *abstract first* (Table 6). Each noun repeats three times in Table 5 and three times in Table 6, six overall. No noun pair repeats. The *concrete-abstract* subcategory pairing repeats once overall (*across* Tables 5 and 6) but not *within* Tables 5 or 6.

We also wanted to make sure that our items were novel, to control for familiarity effects. Past work has used human raters to assess novelty (Shen and Aisenman, 2008). We used a corpus approach. We queried each prime in the BNC using the search operator <<*s*>>. This counts how often a term appears in the same sentence[4] (in a typical British context), not necessarily side-by-side. We did the same for noun pairs. All primes and noun pairs co-occurred in this way at rates of fewer than 20 per 100 million words. We considered this novelty and moved on.

We had two lists of primes and two lists of noun pairs at this stage. Our next step was to combine these into final stimuli lists. This required us to consider how the lists in-hand related to our broader experimental design.

[4] It's important to note that the BNC spoken and written are formatted/transcribed differently (grammatical sentences versus comma-punctuated utterances). When we say 'appears in the same sentence', we refer to the sentence length value (derived from w-tags per s-unit) common to both the written BNC and the spoken BNC. w-tags are considered comparable to words, and s-units are considered comparable to sentences.

2.4. Fitting lists to the design

We thought about our lists in terms of our variables, planned inferential tests, and other design decisions. We use this section to highlight how our view of language is influenced by design decisions and vice versa.

Our next step was about turning finalised item lists into final stimuli lists. We had 90 *preferred-syntax* primes, 90 *nonpreferred-syntax* primes, 9 *literal-first* noun pairs, and 9 *abstract-first* noun pairs. We wanted to avoid interstimulus priming, specifically one syntactic order biasing the other (Bonini et al., 1999; Frazier et al. 2008). This is especially important here because semantic priming effects in schizophrenia appear specific to formal thought disorder (Pomarol-Clotet et al., 2008). We decided to counterbalance by allocating half of the sample to different item sets. We wanted no single participant to see both syntactic possibilities. This decision had widespread design implications. We needed to present all participants with an equal number of *preferred* and *nonpreferred-syntax* primes. We also needed to present all participants with an equal number of *literal-first* and *abstract-first* noun pairs. We also wanted the *prime-noun* combinations to differ. This was complicated by our need to work mainly with sets of 3, 5, and 9. We first cut our prime lists in half and shuffled them. This involved combining the first 45 *nonpreferred-syntax* primes with the last 45 *preferred-syntax* primes. We did the same in reverse, combining the first 45 *preferred-syntax* primes with the last 45 *nonpreferred-syntax* primes. We then distributed our 18 noun pairs across our newly shuffled prime lists. This gave us two distinct lists (Tables 7 and 8). List A (Table 7) uses *nonpreferred-syntax* primes with *literal-first* nouns (for items 1 to 45) and *preferred-syntax* primes with *abstract-first* nouns (for items 45 to 90). List B (Table 8) uses *preferred-syntax* primes with *literal-first* nouns (for items 1–45) and *nonpreferred-syntax* primes with *abstract-first* nouns (for items 45 to 90). This means that list A and B participants experience the same number of *preferred-* and *nonpreferred-syntax* primes, the same number of *literal-first* and *abstract-first* noun pairs, and the same number of the same words. It also means that list A and B participants do not experience the same *prime-response* pairings and see different words for *preferred-syntax* primes, *nonpreferred-syntax* primes, *literal-first* nouns, and *abstract-first* nouns. This helps us understand whether list-specific features influence the data. Each *sensory* adjective (*visual*) repeats 12 times in Table 7 and 12 times in Table 8, 24 overall. No *sensory* adjective pair/prime (*visual smelly*) repeats. Each *sensory* subcategory (*sight*) repeats 36 times in Table 7 and 36 times in Table 8, 72 overall. Each *sensory* subcategory pairing (*sight-sound*) repeats nine times in Table 7 and nine times in Table 8, 18 overall. Each *noun* (*edge*) repeats 30 times in Table 7 and 30 times in Table 8, 60 overall. Each *noun* pair (*edge consideration*) repeats five times in Table 7 and five times in Table 8, 10 overall. Each *noun* subcategory (*general object/thought, belief*) repeats 90 times in Table 7 and 90 times in Table 8, 180 overall. There is only one *noun* subcategory pairing (*general object-thought, belief*) in Table 7 and Table 8. We wanted to make sure that none of these elements loaded in a given direction, such as having 10 *sight* adjectives but only 8 *sound* adjectives. The actual counts are not important to this aim, only equal weightings. We discuss related limitations in Section 3.

We also remembered around this time a forgotten plan to create practice and distractor item lists. We constructed these with the items previously excluded from the main sets. This allowed us to use all generated permutations in a balanced way, except six *noun* subcategory pairings that we discuss later. We draw reader attention to the permutation calculations described in Section 2.3, specifically that these were conducted *without repetition*. We recalculated permutations for the *sensory* adjectives, this time *with repetition*. This gave us 25 permutations. These five extra permutations were *sensory* subcategory duplicates (*touch-touch*). It was possible to create two types of prime within this subcategory. We could repeat *sensory* adjectives (*visual*) or match different *sensory* adjectives from the same subcategory (*visual watchful*). Both types were not used in the main item lists. We also recalculated permutations for the *noun* pairs, this time *with repetition*. This gave us 36 permutations. These six extra permutations were *single-subcategory* noun pairs (*edge-sharpener*). It was also possible to create two types of noun pair within this subcategory. We could repeat *nouns* (*edge*) or match different nouns from the same subcategory (*edge sharpener*). Both types were not used in the main item lists. We assigned duplicate primes (*visual*) to the practice lists because we wanted to signal that they were not proper task items. We assigned same-subcategory (*visual watchful*) primes to the distractor list because they more closely resembled the main items. We then distributed the unused single-subcategory (*boomerang edge*) noun pairs equally across the practice and distractor lists. We had to cut six *duplicates* of the same noun (*edge*) to make Sections 2.2 to 2.4 work. Luckily the only items we could not place were those we did not need. This gave us two practice lists (Tables 9 and 10) and two distractor lists (Tables 11 and 12).

It was not possible to balance *concrete* and *abstract* nouns *within* these lists. This is because we were using *single-subcategory* noun pairs (*edge sharpener*) and working with sets of 5 and 3. We responded by counterbalancing them *across* lists. List A practice items (Table 9) used *concrete* noun pairs, and List B practice items (Table 10) used *abstract* noun pairs. List A distractor items (Table 11) used *abstract* noun pairs, and List B distractor items (Table 12) used *concrete* noun pairs. Practice lists A and B contain the same *sensory* adjective duplicates (*smelly smell*). Distractor lists A and B contain *single-subcategory* primes (*smelly fragrant*). Practice lists A and B use different noun subcategories (*concrete* in A is *abstract* in B). Distractor lists A and B use different noun subcategories (*abstract* in A is *concrete* in B). Distractor lists A and B present primes in different syntactic orders (*fragrant aromatic* in A is *aromatic fragrant* in B). Each *sensory* adjective (*visual*) repeats twice in Table 9, twice in Table 10, twice in Table 11, and twice in Table 12, eight overall. *Sensory* duplicate primes (*aromatic*) repeat *across* Tables 9 and 10 (but not *within* Table 9 or Table 10). No *single-subcategory* prime (*aromatic fragrant*) repeats in Table 11 or Table 12. Each *sensory* subcategory (*sight*) repeats six times in Table 9, six times in Table 10, six times in Table 11, and six times in Table 12, 24 overall. Each *sensory subcategory* pairing (*sound-sound*) repeats three times in Table 9, three times in Table 10, three times in Table 11, and three times in Table 12, 12 overall. Each *sensory single-subcategory* prime (*aromatic fragrant*) repeats once (in reverse) *across* Tables 11 and 12 (but not *within* Table 11 or Table 12). Each *concrete noun* (*edge*) repeats 10 times in Table 9 and 10 times in Table 12, 20 overall. Each *abstract noun* (opinion) repeats 10 times in Table 10 and 10 times in Table 11, 20 overall. Each *concrete noun* pair (*edge sharpener*) repeats five times in Table 9 and five times in Table 12, 10 overall. Each *abstract noun* pair repeats five times in Table 10 and five times in Table 11, 10 overall. The *general object-general object subcategory* pairing repeats once *across* Tables 9 and 12 (but not *within* Tables 9 and 12). The *thought, belief-thought, belief subcategory* pairing repeats once *across* Tables 10 and 11 (but not *within* Tables 10 and 11). This gave us two stimuli pathways, A and B. Each has its own main, practice, and distractor list (see Appendix I). These lists influenced much of the broader design.

The steps in Sections 2.2 to 2.4 led to an important theoretical question: are syntactic variations (*aromatic fragrant* vs *fragrant aromatic*) counterbalanced forms of one experimental unit or two different units? The answer determines whether a repeated measures or independent groups design is used. We answered *linguistically* (two different units) at this point because syntactic differences impact the semantic properties of words and phrases. We nonetheless tested our item lists against several designs (repeated measures and independent groups), to determine best fit and independent groups compatibility. We ruled out a 2 × 2 fully-within subjects design (when there are two independent variables, each with two levels, and all participants experience all items/orders) because it did not accommodate control of interstimulus priming and lengthened the experiment. Complete randomisation required over 2000 presentations for example. We ruled out a fully-between-subjects design (when there are two independent variables, each with two levels, and one group experiences one set of items/order but not the other) because the only way to preserve balanced exposures and control interstimulus priming was by unintentionally pairing stimuli (*preferred syntax* primes with only *concrete-first* nouns). We could not use a 2 × 2 mixed-factorial design (when there are two independent variables, each with two levels, and one of is tested as between-subjects while the other is tested as within-subjects) because it did not accommodate balanced exposures. We then explored suitable inferential tests that accommodated independent groups *and* our assigning half of each population (clinical and control) to different lists. It became clear that we needed a mixed-model analysis to account for missing data (see Goos and Gilmour, 2012; Padilla and Algina, 2007; Næs et al., 2007; Bates et al., 2007). The problem was that our aim was about studying syntactic variation. We also expected exposure to both variations to bias the data. Our lists deal with this by making sure that no participant sees both orders. The differences between both orders are however what we are studying. We therefore need to compare responses to both orders in samples that experienced only one. A traditional crossover is not a solution because we have two populations. We also wanted to make participation comfortable, especially for a clinical cohort with known language difficulties. The chosen design needed to preserve equal exposures, control interstimulus priming, permit randomisation at multiple levels and more without lengthening the experiment. We required a mixed model for all of these reasons. Multi-level and mixed-effects models are incidentally neglected in corpus linguistics (Gries, 2015).

Figs. 1-2

A split-plot design (Fisher, 1960) met all criteria. Split plots are underused in behavioural research (Zhao et al., 2018) and used unknowingly elsewhere (Jones and Nachtsheim, 2009) They suit factors that are hard to vary, like our equalised but differing adjective, noun, prime, and response counts. Factors that are easier to randomise are nested within factors that are harder to randomise. This is ideal because our list structures mirror split-plot structures. We nested noun stimuli within the *sensory subcategory* pairings because the former were easier to randomise. This is why we shifted from calculating permutations of *sensory subcategory* pairings to permutations of specific noun pairs. We used a split-plot to control interstimulus priming for both factors, balance exposures, and counterbalance unintentional pairings without lengthening runs. Fig. 3 (see Appendix II) shows the relationship between participant groups, their list pathways, stimuli lists, and list contents. We draw attention to the fact that this split-plot approach only works with the main item lists. The practice and distractor items cannot be included in this analysis. We consider this acceptable given the function of practice and distractor items.

We then computerised the task. This stage gave us the opportunity to implement additional controls. We used a 950 ms, blank-screen interstimulus interval. We also avoided fixation crosses because of their semiotic properties. Crosses are sometimes used to denote *and*. We wanted to avoid any implied associations between presentations, particularly because this may appeal to overinclusive thinking in schizophrenia. We kept wording, font, font size, and spacing relative to the centre line consistent throughout. We decided on mouse and touchscreen inputs. Participants respond by clicking or touching their choice or its surrounding area. We wanted the response process to be as intuitive and natural as possible due to our interest in response times. We randomised items at several levels. The main and distractor lists were randomised. This meant that the presentation order for main and distractor items was random. We randomised items within each list, per experimental run. This meant that each time a list (main or distractor) was randomly selected, an item from within that list was randomly presented. This randomisation pattern changed with each participant. We block randomised list allocations. This means that every participant was randomly assigned to pathway A or B.

Matching on linguistic ability is important in schizophrenia research (Heinrichs and Zakzanis, 1998). This is usually done with verbal IQ subscales or non-verbal reasoning tasks (Chatzidamianos et al., 2018). We did it with the Cognitive Linguistic Quick Test (CLQT; Helm-Estabrooks, 2001). This test uses subtasks that ask participants to generate items beginning with a specified letter or within a semantic category. Few researchers attempt to control for the priming effects of these assessments. We expected asking participants to repeat *m* words for one minute and related tasks to have some influence on their task performance in our experiment. We therefore block randomised the order of our screening measure. This means participants sat their CLQT(+) subtasks either before or after the experiment on random assignment. We did this to counterbalance any priming effects caused by the CLQT(+). Visual overviews of our list allocations and block randomisations are under Figs. 3 and 4 (see Appendix II).

3. Discussion

The development process was a learning experience and highlighted a lack of work in the area. We hope that we have shown the value of corpus-driven norming approaches. We also hope that readers find our solutions to the challenges encountered interesting if not helpful. We would be pleased if others used our materials.

Our design of course has limitations. It for one does not allow us to tell whether the preferences expressed by participants are semantic or pragmatic in nature. This could be explored by comparing task performance across groups with and without known pragmatic disabilities. Although we controlled for interstimulus priming and made sure that items occurred at equal rates within their sets, we did not resolve the problem of differences in set sizes, such as noun pairings appearing more often than adjective pairings. We attempted several stimuli arrangements to avoid this. The one presented here is the best result from those attempts. We expect that this problem goes away when the pool of candidate words is large enough. We had to weigh the balance of having fewer but better controlled candidate words over the opposite. The fact that *sensory* adjectives are sensitive to multiple psycholinguistic confounders also had a lot to do with this, creating a need for multiple control stages that will not be necessary for all studies. We expect that this type of issue is more likely to affect scalar stimuli than function words or common words with better inherent construct validity.

We also recognise that the BNC may not reflect language in schizophrenia. Our view is supported by work that compares corpus and within-sample word frequency approaches. The within-sample approach predicted positive symptoms, and the corpus did not (Gabrić et al. 2021). This makes sense. A corpus of speech in schizophrenia should offer better symptom prediction than one built to represent an entire language variety. We used the BNC because (1) we had no specialised corpus and (2) our clinical participants experience British English. We nonetheless understand the need for a schizophrenia corpus. This is why we interviewed our experimental participants and built one. We describe that corpus and its construction in another paper.

Our main takeaways are that designing language experiments is easier when experimental units are defined (1) early on and (2) linguistically.

Funding

this work is part of a self-funded PhD.

Declaration of Competing Interest

We declare no competing interests.

Appendix I

Tables 1-12

Table 1
Sensory adjectives by *semantic* subcategory.

Sight (X3.4)	**sound (X3.2)**	**smell (X3.5)**	**taste (X3.1)**	**touch (X3.3)**
visual	quiet	aromatic	sour	rough
watchful	silent	fragrant	spicy	smoothed
scanned	deaf	smelly	salty	stroked

Table 2
Semantically sensory adjectives, CCT preferred syntax.

Smell-sight	**sound-sight**	**taste-sight**	**touch-sight**	**taste-smell**
aromatic visual	quiet visual	sour visual	rough visual	sour aromatic
aromatic watchful	quiet watchful	sour watchful	rough watchful	sour fragrant
aromatic scanned	quiet scanned	sour scanned	rough scanned	sour smelly
fragrant visual	silent visual	spicy visual	smoothed visual	spicy aromatic
fragrant watchful	silent watchful	spicy watchful	smoothed watchful	spicy fragrant
fragrant scanned	silent scanned	spicy scanned	smoothed scanned	spicy smelly
smelly visual	deaf visual	salty visual	stroked visual	salty aromatic
smelly watchful	deaf watchful	salty watchful	stroked watchful	salty fragrant
smelly scanned	deaf scanned	salty scanned	stroked scanned	salty smelly
touch-smell	**smell-sound**	**taste-sound**	**touch-sound**	**touch-taste**
rough aromatic	aromatic quiet	sour quiet	rough quiet	rough sour
rough fragrant	aromatic silent	sour silent	rough silent	rough spicy
rough smelly	aromatic deaf	sour deaf	rough deaf	rough salty
smoothed aromatic	fragrant quiet	spicy quiet	smoothed quiet	smoothed sour
smoothed fragrant	fragrant silent	spicy silent	smoothed silent	smoothed spicy
smoothed smelly	fragrant deaf	spicy deaf	smoothed deaf	smoothed salty
stroked aromatic	smelly quiet	salty quiet	stroked quiet	stroked sour
stroked fragrant	smelly silent	salty silent	stroked silent	stroked spicy
stroked smelly	smelly deaf	salty deaf	stroked deaf	stroked salty

Table 3
Semantically sensory adjectives, CCT nonpreferred syntax.

Sight-smell	sight-sound	sight-taste	sight-touch	smell-taste
visual aromatic	visual quiet	visual sour	visual rough	aromatic sour
watchful aromatic	watchful quiet	watchful sour	watchful rough	fragrant sour
scanned aromatic	scanned quiet	scanned sour	scanned rough	smelly sour
visual fragrant	visual silent	visual spicy	visual smoothed	aromatic spicy
watchful fragrant	watchful silent	watchful spicy	watchful smoothed	fragrant spicy
scanned fragrant	scanned silent	scanned spicy	scanned smoothed	smelly spicy
visual smelly	visual deaf	visual salty	visual stroked	aromatic salty
watchful smelly	watchful deaf	watchful salty	watchful stroked	fragrant salty
scanned smelly	scanned deaf	scanned salty	scanned stroked	smelly salty
smell-touch	**sound-smell**	**sound-taste**	**sound-touch**	**taste-touch**
aromatic rough	quiet aromatic	quiet sour	quiet rough	sour rough
fragrant rough	silent aromatic	silent sour	silent rough	spicy rough
smelly rough	deaf aromatic	deaf sour	deaf rough	salty rough
aromatic smoothed	quiet fragrant	quiet spicy	quiet smoothed	sour smoothed
fragrant smoothed	silent fragrant	silent spicy	silent smoothed	spicy smoothed
smelly smoothed	deaf fragrant	deaf spicy	deaf smoothed	salty smoothed
aromatic stroked	quiet smelly	quiet salty	quiet stroked	sour stroked
fragrant stroked	silent smelly	silent salty	silent stroked	spicy stroked
smelly stroked	deaf smelly	deaf salty	deaf stroked	salty stroked

Table 4
Semantically concrete and abstract nouns.

Concrete (O2: general object)	abstract (X2.1: thought, belief)
edge	opinion
sharpener	attitude
boomerang	consideration

Table 5
Concrete-first noun pairs.

Concrete first (O2-X2.1)
edge opinion
edge attitude
edge consideration
sharpener opinion
sharpener attitude
sharpener consideration
boomerang opinion
boomerang attitude
boomerang consideration

Table 6
Abstract-first noun pairs.

Abstract first (X2.1-O2)
opinion edge
opinion sharpener
opinion boomerang
attitude edge
attitude sharpener
attitude boomerang
consideration edge
consideration sharpener
consideration boomerang

Table 7
Pathway A, items 1 to 45: *nonpreferred-syntax* primes, *literal-first* nouns; items 45 to 90: *preferred-syntax* primes, *abstract-first* nouns.

Prime	response
visual aromatic	edge opinion
watchful aromatic	edge attitude
scanned aromatic	edge consideration
visual fragrant	sharpener opinion
watchful fragrant	sharpener attitude
scanned fragrant	sharpener consideration
visual smelly	boomerang opinion
watchful smelly	boomerang attitude
scanned smelly	boomerang consideration
visual quiet	edge opinion
watchful quiet	edge attitude
scanned quiet	edge consideration
visual silent	sharpener opinion
watchful silent	sharpener attitude
scanned silent	sharpener consideration
visual deaf	boomerang opinion
watchful deaf	boomerang attitude
scanned deaf	boomerang consideration
visual sour	edge opinion
watchful sour	edge attitude
scanned sour	edge consideration
visual spicy	sharpener opinion
watchful spicy	sharpener attitude
scanned spicy	sharpener consideration
visual salty	boomerang opinion
watchful salty	boomerang attitude
scanned salty	boomerang consideration
visual rough	edge opinion
watchful rough	edge attitude
scanned rough	edge consideration
visual smoothed	sharpener opinion
watchful smoothed	sharpener attitude
scanned smoothed	sharpener consideration

Table 7
(continued)

Prime	response
visual stroked	boomerang opinion
watchful stroked	boomerang attitude
scanned stroked	boomerang consideration
aromatic sour	edge opinion
fragrant sour	edge attitude
smelly sour	edge consideration
aromatic spicy	sharpener opinion
fragrant spicy	sharpener attitude
smelly spicy	sharpener consideration
aromatic salty	boomerang opinion
fragrant salty	boomerang attitude
smelly salty	boomerang consideration
rough aromatic	opinion edge
rough fragrant	opinion sharpener
rough smelly	opinion boomerang
smoothed aromatic	attitude edge
smoothed fragrant	attitude sharpener
smoothed smelly	attitude boomerang
stroked aromatic	consideration edge
stroked fragrant	consideration sharpener
stroked smelly	consideration boomerang
aromatic quiet	opinion edge
aromatic silent	opinion sharpener
aromatic deaf	opinion boomerang
fragrant quiet	attitude edge
fragrant silent	attitude sharpener
fragrant deaf	attitude boomerang
smelly quiet	consideration edge
smelly silent	consideration sharpener
smelly deaf	consideration boomerang
sour quiet	opinion edge
sour silent	opinion sharpener
sour deaf	opinion boomerang
spicy quiet	attitude edge
spicy silent	attitude sharpener
spicy deaf	attitude boomerang
salty quiet	consideration edge
salty silent	consideration sharpener
salty deaf	consideration boomerang
rough quiet	opinion edge
rough silent	opinion sharpener
rough deaf	opinion boomerang
smoothed quiet	attitude edge
smoothed silent	attitude sharpener
smoothed deaf	attitude boomerang
stroked quiet	consideration edge
stroked silent	consideration sharpener
stroked deaf	consideration boomerang
rough sour	opinion edge
rough spicy	opinion sharpener
rough salty	opinion boomerang
smoothed sour	attitude edge
smoothed spicy	attitude sharpener
smoothed salty	attitude boomerang
stroked sour	consideration edge
stroked spicy	consideration sharpener
stroked salty	consideration boomerang

Table 8
Pathway B, items 1 to 45: *preferred-syntax* primes, *literal-first* nouns; items 45 to 90: *nonpreferred-syntax* primes, *abstract-first* nouns.

Prime	response
aromatic visual	edge opinion
aromatic watchful	edge attitude
aromatic scanned	edge consideration
fragrant visual	sharpener opinion
fragrant watchful	sharpener attitude
fragrant scanned	sharpener consideration
smelly visual	boomerang opinion
smelly watchful	boomerang attitude
smelly scanned	boomerang consideration
quiet visual	edge opinion
quiet watchful	edge attitude
quiet scanned	edge consideration
silent visual	sharpener opinion
silent watchful	sharpener attitude
silent scanned	sharpener consideration
deaf visual	boomerang opinion
deaf watchful	boomerang attitude
deaf scanned	boomerang consideration
sour visual	edge opinion
sour watchful	edge attitude
sour scanned	edge consideration
spicy visual	sharpener opinion
spicy watchful	sharpener attitude
spicy scanned	sharpener consideration
salty visual	boomerang opinion
salty watchful	boomerang attitude
salty scanned	boomerang consideration
rough visual	edge opinion
rough watchful	edge attitude
rough scanned	edge consideration
smoothed visual	sharpener opinion
smoothed watchful	sharpener attitude
smoothed scanned	sharpener consideration
stroked visual	boomerang opinion
stroked watchful	boomerang attitude
stroked scanned	boomerang consideration
sour aromatic	edge opinion
sour fragrant	edge attitude
sour smelly	edge consideration
spicy aromatic	sharpener opinion
spicy fragrant	sharpener attitude
spicy smelly	sharpener consideration
salty aromatic	boomerang opinion
salty fragrant	boomerang attitude
salty smelly	boomerang consideration
aromatic rough	opinion edge
fragrant rough	opinion sharpener
smelly rough	opinion boomerang
aromatic smoothed	attitude edge
fragrant smoothed	attitude sharpener
smelly smoothed	attitude boomerang
aromatic stroked	consideration edge
fragrant stroked	consideration sharpener
smelly stroked	consideration boomerang
quiet aromatic	opinion edge
silent aromatic	opinion sharpener
deaf aromatic	opinion boomerang
quiet fragrant	attitude edge
silent fragrant	attitude sharpener
deaf fragrant	attitude boomerang
quiet smelly	consideration edge
silent smelly	consideration sharpener
deaf smelly	consideration boomerang
quiet sour	opinion edge
silent sour	opinion sharpener
deaf sour	opinion boomerang
quiet spicy	attitude edge
silent spicy	attitude sharpener
deaf spicy	attitude boomerang
quiet salty	consideration edge
silent salty	consideration sharpener
deaf salty	consideration boomerang
quiet rough	opinion edge
silent rough	opinion sharpener

Table 8
(continued)

Prime	response
deaf rough	
quiet smoothed	opinion boomerang
silent smoothed	attitude edge
deaf smoothed	attitude sharpener
quiet stroked	attitude boomerang
silent stroked	consideration edge
deaf stroked	consideration sharpener
sour rough	consideration boomerang
spicy rough	opinion edge
salty rough	opinion sharpener
sour smoothed	opinion boomerang
spicy smoothed	attitude edge
salty smoothed	attitude sharpener
sour stroked	attitude boomerang
spicy stroked	consideration edge
salty stroked	consideration sharpener
	consideration boomerang

Table 9
Practice List A.

Prime	response
aromatic	sharpener edge
fragrant	boomerang edge
smelly	boomerang sharpener
sour	sharpener edge
spicy	boomerang edge
salty	boomerang sharpener
visual	sharpener edge
watchful	boomerang edge
scanned	boomerang sharpener
quiet	sharpener edge
silent	boomerang edge
deaf	boomerang sharpener
rough	sharpener edge
smoothed	boomerang edge
stroked	boomerang sharpener

Table 10
Practice List B.

Prime	response
aromatic	attitude opinion
fragrant	consideration opinion
smelly	consideration attitude
sour	attitude opinion
spicy	consideration opinion
salty	consideration attitude
visual	attitude opinion
watchful	consideration opinion
scanned	consideration attitude
quiet	attitude opinion
silent	consideration opinion
deaf	consideration attitude
rough	attitude opinion
smoothed	consideration opinion
stroked	consideration attitude

Table 11
Distractor List A,.

Prime	response
fragrant aromatic	attitude opinion
smelly aromatic	consideration opinion
smelly fragrant	consideration attitude
spicy sour	attitude opinion
salty sour	consideration opinion
salty spicy	consideration attitude
watchful visual	attitude opinion
scanned visual	consideration opinion
scanned watchful	consideration attitude
silent quiet	attitude opinion
deaf quiet	consideration opinion
deaf silent	consideration attitude
smoothed rough	attitude opinion
stroked rough	consideration opinion
stroked smoothed	consideration attitude

Table 12
Distractor List B.

Prime	response
aromatic fragrant	sharpener edge
aromatic smelly	boomerang edge
fragrant smelly	boomerang sharpener
sour spicy	sharpener edge
sour salty	boomerang edge
spicy salty	boomerang sharpener
visual watchful	sharpener edge
visual scanned	boomerang edge
watchful scanned	boomerang sharpener
quiet silent	sharpener edge
quiet deaf	boomerang edge
silent deaf	boomerang sharpener
rough smoothed	sharpener edge
rough stroked	boomerang edge
smoothed stroked	boomerang sharpener

Appendix II

Figs. 1-4

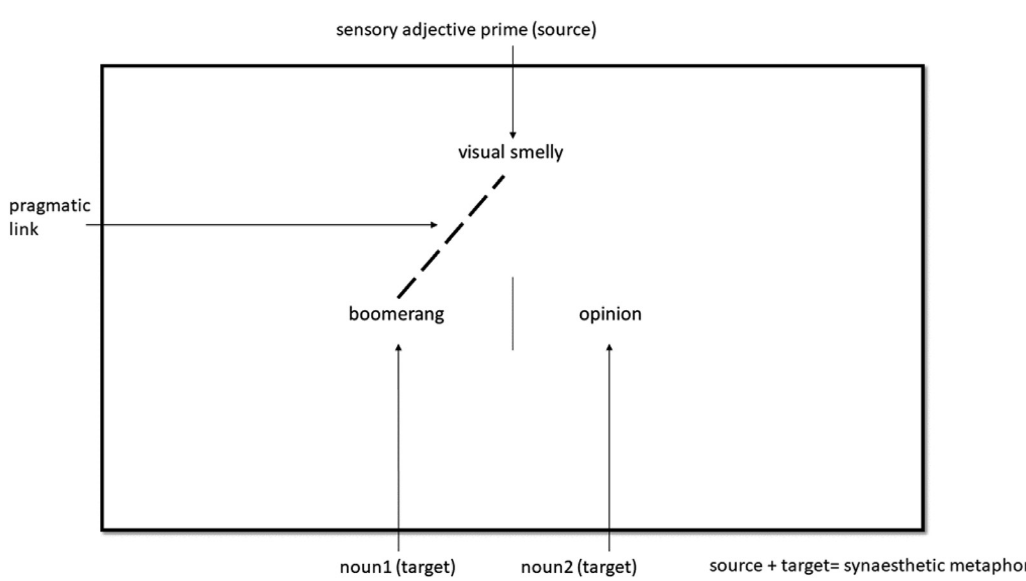

Fig. 1. Example presentation, annotated for the reader.

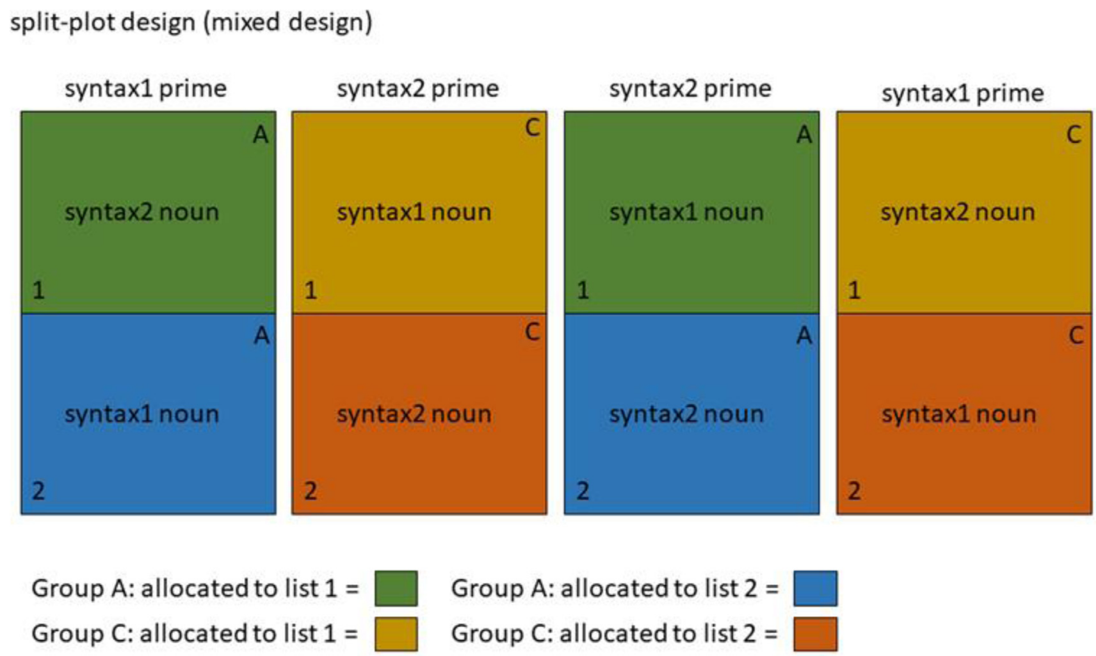

Fig. 2. Groups, pathways, main item lists, and list contents in split-plot
Note
syntax1 prime = CCT preferred syntax prime
syntax2 prime = CCT nonpreferred syntax prime
syntax1 noun = concrete-first noun pair
syntax2 noun = abstract-first noun pair
Group *A* = clinical cohort
Group C = comparison cohort
List 1 = pathway A (List A, Practice List A, Distractor List A)
List 2 = pathway B (List B, Practice List B, Distractor List B).

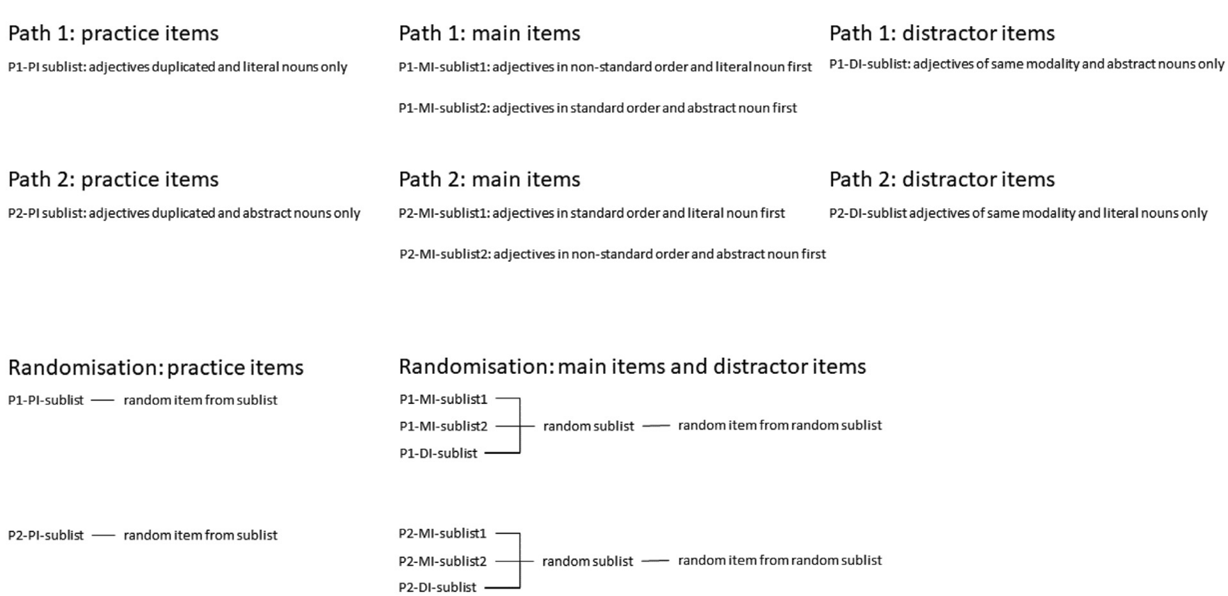

Fig. 3. List contents and randomisation.

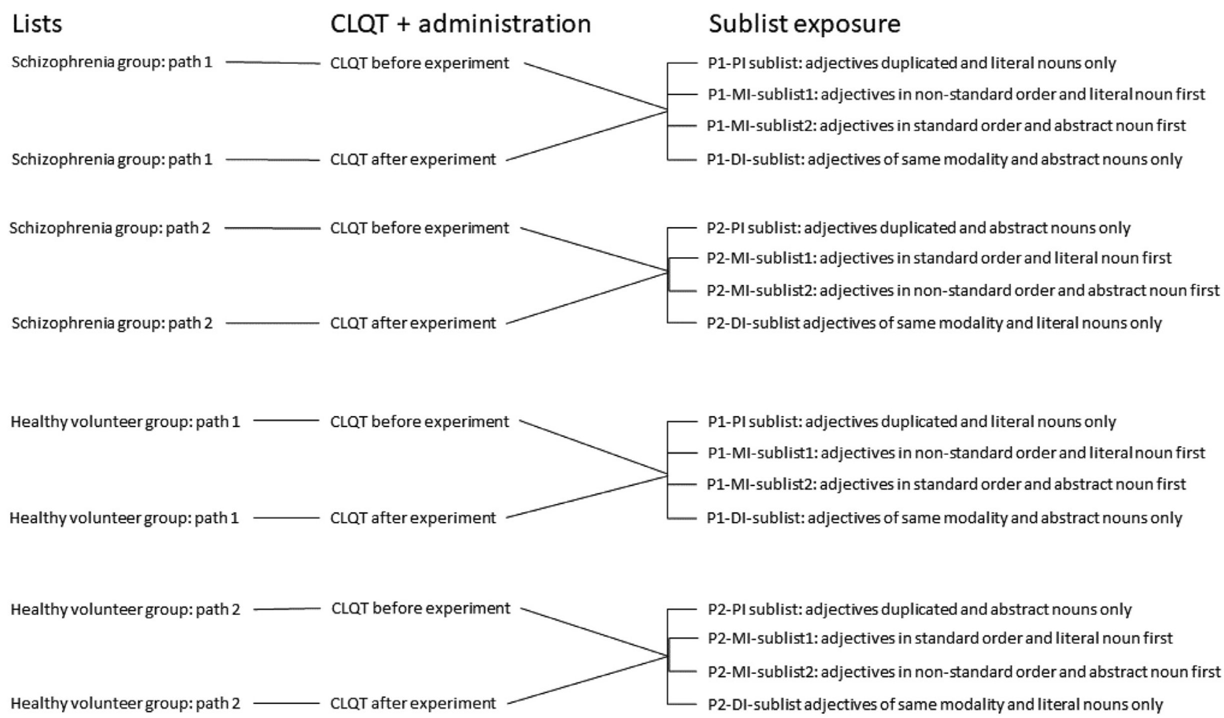

Fig. 4. Group, list, assessment, and sublist block randomisations.

References

Bartsch, K., Wellman, H.M., 1995. Children Talk About the Mind. Oxford university press.

Bates, D., Sarkar, D., Bates, M.D., Matrix, L., 2007. The lme4 package. R Package Version 2, 74.

Bezuidenhout, A., 2010. Experimental pragmatics. In: The Pragmatics Encyclopedia, pp. 148–153.

Bilgrami, Z., Gutierrez, E., Sarac, C., Corcoran, C., & Cecchi, G. (2020). S108. Automated detection of increased metaphor production in open-ended speech in schizophrenia and its risk states. Schizophr Bull., 46(Suppl 1), S75. https://doi.org/10.1093/schbul/sbaa031.174

Bomarius, F., Oivo, M., Jaring, P., Abrahamsson, P.eds., 2009. *Product-Focused Software Process Improvement: 10th International ConferencePROFES 2009, Oulu, Finland, June 15-17, 2009, Proceedings* (Vol. 32). Springer Science & Business Media.

Bonini, N., Osherson, D., Viale, R., Williamson, T., 1999. On the psychology of vague predicates. Mind Lang. 14 (4), 377–393. doi:10.1111/1468-0017.00117.

Brennan, S.E., 1995. Centering attention in discourse. Lang. Cogn. Process. 10 (2), 137–167. doi:10.1080/01690969508407091.

Burnard, L., 2002. Where did we go wrong? A retrospective look at the British National Corpus. In: Teaching and Learning By Doing Corpus Analysis. Brill Rodopi, pp. 51–70. doi:10.1163/9789004334236_007.

Caspers, J., 2003. Local speech melody as a limiting factor in the turn-taking system in Dutch. J. Phon. 31 (2), 251–276. doi:10.1016/S0095-4470(03)00007-X.

Chatzidamianos, G., McCarthy, R.A., Du Feu, M., Rosselló, J., McKenna, P.J., 2018. Language abnormality in deaf people with schizophrenia: a problem with classifiers. Cogn. Neuropsychiatry 23 (4), 229–241.

Chen, X., Meurers, D., 2018. Word frequency and readability: predicting the text-level readability with a lexical-level attribute. J. Res. Read. 41 (3), 486–510.

Clark, H.H., Bangerter, A., 2004. Changing ideas about reference. In: Experimental Pragmatics. Palgrave Macmillan, London, pp. 25–49. doi:10.1057/9780230524125_2.

Degen, J., 2015. Investigating the distribution of some (but not all) implicatures using corpora and web-based methods. Semant. Pragmat. 8, 11. doi:10.3765/sp.8.11, -1.

Dewhurst, S.A., Hitch, G.J., Barry, C., 1998. Separate effects of word frequency and age of acquisition in recognition and recall. J. Experiment. Psychol. 24 (2), 284.

Eiteljoerge, S.F., Pouscoulous, N., Lieven, E.V., 2018. Some pieces are missing: implicature production in children. Front. Psychol. 9, 1928. doi:10.3389/fpsyg.2018.01928.

Ferber, R. (1995). Reliability and validity of slip-of-the-tongue corpora: a methodological note. https://doi.org/10.1515/ling.1995.33.6.1169

Fernández, E.M., Cairns, H.S., Wiley, J., 2018. The Handbook of Psycholinguistics. John Wiley & Sons Incorporated.

Fisher, R.A., 1960. The Design of experiments. The design of Experiments, 7th Ed..

Frazier, L., 2006a. The Big Fish in a Small pond: Accommodation and the Processing of Novel Definites October. Ohio State University Presupposition Accommodation Workshop, Columbus, Ohio.

Garnham, A., Shillcock, R.C., Brown, G.D., Mill, A.I., Cutler, A., 1982. Slips of the tongue in the London-Lund corpus of spontaneous conversation. In: Slips of the Tongue and Language Production, pp. 251–263. doi:10.1515/9783110828306.251.

Gibbs, R.J., Ortony, A., 2008. Metaphor and thought. In: The Cambridge Handbook of Metaphor and Thought, p. 3.

Gilquin, G. and Gries, S.T., 2009. Corpora and experimental methods: a state-of-the-art review.

Goos, P., Gilmour, S.G., 2012. A general strategy for analyzing data from split-plot and multistratum experimental designs. Technometrics 54 (4), 340–354. doi:10.1080/00401706.2012.694777.

Heinrichs, R.W., Zakzanis, K.K., 1998. Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. Neuropsychology 12 (3), 426. doi:10.1037/0894-4105.12.3.426.

Heldner, M., Edlund, J., 2010. Pauses, gaps and overlaps in conversations. J. Phon. 38 (4), 555–568. doi:10.1016/j.wocn.2010.08.002.

Holyoak, K.J., Stamenković, D., 2018. Metaphor comprehension: a critical review of theories and evidence. Psychol. Bull. 144 (6), 641.

Huisman, J.L., Majid, A., 2018. Psycholinguistic variables matter in odor naming. Mem. Cognit. 46 (4), 577–588.

Hwang, Y.T., Strikwerda-Brown, C., El-Omar, H., Ramanan, S., Hodges, J.R., Burrell, J.R., Piguet, O., Irish, M., 2021. More than words"–Longitudinal linguistic changes in the works of a writer diagnosed with semantic dementia. Neurocase 1–10.

Isaacs, T., Murdoch, J., Demjén, Z., Stevenson, F., 2020. Examining the language demands of informed consent documents in patient recruitment to cancer trials using tools from corpus and computational linguistics. Health doi:10.1177/1363459320963431, 1363459320963431.

Jones, B., Nachtsheim, C.J., 2009. Split-plot designs: what, why, and how. J. Qual. Technol. 41 (4), 340–361. doi:10.1080/00224065.2009.11917790.

Lakoff, G., Johnson, M., 1980. Conceptual metaphor in everyday language. J. Philos. 77 (8), 453–486.

Liao, S.Y., Meskin, A., 2017. Aesthetic adjectives: experimental semantics and context-sensitivity. Philos Phenomenol Res 94 (2), 371–398. doi:10.1111/phpr.12217.

Liao, S.Y., McNally, L., Meskin, A., 2016. Aesthetic adjectives lack uniform behavior. Inquiry 59 (6), 618–631. doi:10.1080/0020174X.2016.1208927.

Mandera, P., Keuleers, E., Brysbaert, M., 2015. How useful are corpus-based methods for extrapolating psycholinguistic variables? Q. J. Exp. Psychol. 68 (8), 1623–1642.

McKenna, P.J., Oh, T.M., 2005. Schizophrenic speech: Making sense of Bathroots and Ponds That Fall in Doorways. Cambridge University Press.

Munday, I., Newton-John, T., Kneebone, I., 2020. Barbed wire wrapped around my feet': metaphor use in chronic pain. Br. J. Health Psychol. e12432. doi:10.1111/bjhp.12432.

Næs, T., Aastveit, A.H., Sahni, N.S., 2007. Analysis of split-plot designs: an overview and comparison of methods. Qual. Reliab. Eng. Int. 23 (7), 801–820. doi:10.1002/qre.841.

Nation, I.S.P., 2004. A study of the most frequent word families in the British National Corpus. In: Vocabulary in a Second Language: Selection, Acquisition and Testing. John Benjamins, Amsterdam, pp. 3–13.

Noveck, I.A., Sperber, D. (Eds.), 2004. Experimental Pragmatics. Palgrave Macmillan, Basingstoke doi:10.1057/9780230524125.

Padilla, M.A., Algina, J., 2007. Type I error rates of the Kenward-Roger adjusted degree of freedom F-test for a split-plot design with missing values. J. Modern Appl. Statistic. Methods 6 (1), 8. doi:10.22237/jmasm/1177992420.

Pagel, M., Meade, A., 2018. The deep history of the number words. Philosoph. Trans. R. Soc. B 373 (1740), 20160517.
Pennebaker, J.W., Boyd, R.L., Jordan, K. and Blackburn, K., 2015. *The development and psychometric properties of LIWC2015*.
Perret, C., Bonin, P., 2019. Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. Behav. Res. Methods 51 (6), 2533–2545. doi:10.3758/s13428-018-1100-1.
Pomarol-Clotet, E., Oh, T.M.S.S., Laws, K.R., McKenna, P.J., 2008. Semantic priming in schizophrenia: systematic review and meta-analysis. Br. J. Psychiatry 192 (2), 92–97. doi:10.1192/bjp.bp.106.032102.
Rayson, P., 2008. From key words to key semantic domains. Int. J. Corpus Linguist. 13 (4), 519–549.
Rayson, P., Archer, D., Piao, S. and McEnery, A.M., 2004. The UCREL semantic analysis system.
Sassoon, G.W., 2012. The double nature of negative antonymy. Proc. Sinn und Bedeutung 16 (2), 543–556.
Sassoon, G.W., 2013. A typology of multidimensional adjectives. J. Semant. 30 (3), 335–380. doi:10.1093/jos/ffs012.
Schreuder, R., Kerkman, H., 1987. On the Use of a Lexical Database. In: Psycholinguistic Research in Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora. Costerus, p. 59.
Schwartz, M.F., Saffran, E.M., Bloch, D.E., Dell, G.S., 1994. Disordered speech production in aphasic and normal speakers. Brain Lang. 47 (1), 52–88. doi:10.1006/brln.1994.1042.
Shen, Y., Aisenman, R., 2008. Heard melodies are sweet, but those unheard are sweeter': synaesthetic metaphors and cognition. Lang. Literature 17 (2), 107–121.
Shen, Y., 2002. Cognitive constraints on verbal creativity. Cognit. Stylistics 1, 211.
Stevenson, R.J., Crawley, R.A., Kleinman, D., 1994. Thematic roles, focus and the representation of events. Lang Cogn Process 9 (4), 519–548. doi:10.1080/01690969408402130.
Th Gries, S., 2015. The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models. Corpora 10 (1), 95–125.
Tiemann, S., Schmid, M., Bade, N., Rolke, B., Hertrich, I., Ackermann, H., … Beck, S., 2011. Psycholinguistic evidence for presuppositions: on-line and off-line data. Proc. Sinn und Bedeutung 15, 581–596.
Van Der Henst, J.B., Sperber, D., 2004. Testing the cognitive and communicative principles of relevance. In: Experimental Pragmatics. Palgrave Macmillan, London, pp. 141–171. doi:10.1057/9780230524125_7.
Van Tiel, B., Van Miltenburg, E., Zevakhina, N., Geurts, B., 2016. Scalar diversity. J Semant 33 (1), 137–175.
Westera, M., Brasoveanu, A., 2014. Ignorance in context: the interaction of modified numerals and QUDs. In: Semantics and Linguistic Theory, pp. 414–431. doi:10.3765/salt.v24i0.2436 Vol. 24.
Wijnen, F., 1992. Incidental word and sound errors in young speakers. J. Mem. Lang. 31 (6), 734–755. doi:10.1016/0749-596X(92)90037-X.
Wolters, M., Beaver, D., 2001. What does he mean? In: Proceedings of the Annual Meeting of the Cognitive Science Society Vol. 23, No. 23.
Zhao, A., Ding, P., Mukerjee, R., Dasgupta, T., 2018. Randomization-based causal inference from split-plot designs. Annals Statistics 46 (5), 1876–1903.
Ziemer, K.S., Korkmaz, G., 2017. Using text to predict psychological and physical health: a comparison of human raters and computerized text analysis. Comput. Human Behav. 76, 122–127. doi:10.1016/j.chb.2017.06.038.