Universiteit Antwerpen

This item is the archived peer-reviewed author-version of:

Effects of online abstraction on adjective order preferences

Reference:

Vandekerckhove Bram, Sandra Dominiek, Daelemans Walter.- Effects of online abstraction on adjective order preferences
Language, cognition and neuroscience - ISSN 2327-3798 - 30:7(2015), p. 816-831
Full text (Publishers DOI): http://dx.doi.org/doi:10.1080/23273798.2015.1016977
To cite this reference: http://hdl.handle.net/10067/1236680151162165141

uantwerpen.be

Institutional repository IRUA

Effects of online abstraction on adjective order preferences

Bram Vandekerckhove*

Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK

Tel.: +44 1865 281861

Email: bram.vandekerckhove@psy.ox.ac.uk

Dominiek Sandra

CLiPS Research Center, University of Antwerp, Prinsstraat 13, B-2000 Antwerpen, Belgium

Tel.: +32 3 265 4247

Email: dominiek.sandra@uantwerpen.be

Walter Daelemans

CLiPS Research Center, University of Antwerp, Prinsstraat 13, B-2000 Antwerpen, Belgium

Tel.: +32 3 265 5222

Email: walter.daelemans@uantwerpen.be

* Corresponding author

This research was supported by a PhD fellowship from the Research Foundation – Flanders (FWO), awarded to the first author.

Effects of online abstraction on adjective order preferences

This study investigates whether the generalization of prenominal adjective order preferences is best accounted for by a linear precedence relationship between persistent abstract adjective categories (category precedence) or by online abstraction over adjective exemplars. In a two-alternative forced choice task, participants had to select which of two adjective orders they preferred for low-frequency adjective pairs. Online abstraction over exemplars, operationalized in a range of exemplar-based variables, explained variation in the participants' adjective order choices that category precedence could not account for. However, category precedence still uniquely explained variation in adjective order preference not captured by the exemplar-based variables. Although these findings might support a dual-mechanism model of adjective order preferences, with additive roles for abstract adjective categories and online abstraction, an improved operationalization of online abstraction and a stricter control of the covariates introduced through the implementation of both theoretical approaches are likely to eliminate the effect of category precedence.

Keywords: adjective; word order; exemplar model; categorization; syntax–semantics interface

Introduction

If a language allows sequences of attributive adjectives, there is a universal tendency for those adjectives to occur in a set order (see e.g. Dixon, 1982). In a neutral pragmatic context, any reordering of the adjectives in the noun phrase "the long narrow white Victorian living room" (Adams, 1999, p. 10) makes it sound less natural. Also universal is the finding that those order preferences generalize to groups of semantically and/or

morphologically similar adjectives. Category-based adjective order accounts posit a linear precedence relationship between abstract adjective categories to explain those general order preferences. Dixon (1982), for instance, proposed the following universal order of adjective categories (according to their increasing proximity to the head noun, which means the order is reversed for head-first languages): VALUE ≺ DIMENSION ≺ PHYSICAL PROPERTY ≺ SPEED ≺ HUMAN PROPENSITY ≺ AGE ≺ COLOUR.

A number of researchers have adopted category-based accounts to explain language users' sensitivities to adjective order constraints in language comprehension. In two self-paced reading experiments, Kennison (2010) showed that violations of category-based adjective order constraints caused processing difficulties. Participants' reading times increased at the second adjective in adjective pairs that violated a SIZE ≺ COLOUR constraint (e.g. ... *the red big balloon* ...) (experiment 1) and a GENERAL DESCRIPTION ≺ COLOUR constraint (e.g. ... *the red old box* ...) (experiment 2). Additionally, Kemmerer, Tranel, and Zdanczyk (2009) identified brain-lesioned patients that showed reduced sensitivity to the linear precedence relationship VALUE ≺ SIZE ≺ DIMENSION ≺ PHYSICAL PROPERTY ≺ COLOUR in a two-alternative forced choice (2AFC) task contrasting preferred and dispreferred adjective orders.

Category-based accounts capture adjective order regularities at an abstract level of representation, namely that of the adjectives' positional categories. Still, barring factors external to the adjectives themselves, such as properties of the head noun and the utterance context, and word-level features such as word length and frequency, it remains to be seen whether category precedence explains all non-trivial adjective order variation. In other words, was Dixon (1982) right when he stated that "there appears to be an underlying order between types [i.e. semantic adjective categories such as DIMENSION or COLOUR] but not within types" (p. 25)?

Overall, stable abstract categories are convenient theoretical constructs to describe linguistic generalizations. However, to serve their descriptive purpose, they are often too abstract to capture the full range of non-random variation in the generalization of productive linguistic patterns. Conversely, if they are made concrete enough to capture fine-grained differences between exemplars, they run the danger of becoming indistinguishable from the exemplars they are supposed to generalize over, and hence lose their explanatory value. Exemplar models, on the other hand, are well suited to explain both high- and low-level generalizations. Instead of abstracting away categories from clusters of exemplars on the basis of their shared (semantic or other) features and then accounting for linguistic generalizations in terms of those categories, these models explain linguistic pattern generalizations directly as a result of similarity-based reasoning. The "categories" they employ to abstract away from specific exemplars are therefore not persistent abstract summary representations, but the result of *online abstraction* (Barsalou, 2005). That way, each exemplar determines its own unique temporary category, allowing such models to capture fine-grained generalization behaviour.

A considerable number of studies have shown that theories relying on abstract linguistic categories cannot capture more subtle differences in linguistic behaviour—because the categories are too abstract—or gradient linguistic phenomena—because the categories are too rigid. An example of such a phenomenon in phonology is Bybee's (2001) finding that English double-marked past tenses (i.e. past tenses that are marked by both a vowel change and a dental suffix, such as *left* or *told*) are more likely to be phonetically reduced by deleting the final /t/ or /d/ in high-frequency verbs than in low-frequency verbs. Apart from word effects on phonetic reduction, there are also phrasal effects, suggesting that whole phrases are stored with considerable phonetic detail (for

an overview, see Hay & Bresnan, 2006). In addition, there is evidence that phrase-level factors can affect phonetic change. Hay and Bresnan (2006), for example, showed that a sound change typical of New Zealand English, the raising of the /æ/ vowel (as in *cat* or *sad*), was more likely to be found in the word *hand* when it referred to the limb than when it occurred in phrases such as *give/lend a hand* or *have a hand in*. Whereas category-based phonological theories cannot account for these findings, exemplar models of phonology (e.g. Pierrehumbert, 2001) naturally predict these effects.

Within the domain of inflectional morphology, Ernestus and Baayen (2004) found that the choice between the *-te* and the *-de* allomorphs of the Dutch past tense suffix was not solely determined by the binary [voice]-specification of the stem-final obstruent, contrary to what a category-based account would predict. Instead, this choice was co-determined by the distribution of the stem-final obstruent [voice]-specifications among the phonological neighbours of the verb to be inflected (also see Ernestus & Baayen, 2003; Ernestus, Baayen, van der Wouden, & Broekhuis, 2001). Keuleers et al. (2007) demonstrated that Dutch plural inflection is partly driven by non-phonological analogy. Snider (2007) provided evidence for an exemplar model of syntactic knowledge in production. He showed that priming of relative clause attachment (high vs. low) is modulated by the probabilities of the specific head nouns being modified by a relative clause for both the prime and the target sentences. Snider (2007) also found that ditransitive priming is stronger to the extent that the prime and target sentences are similar in terms of pronominality, animacy, givenness, etc. of the recipient and theme arguments, on top of the "lexical boost" (Pickering & Branigan, 1998).

With regard to comprehenders' sensitivity to adjective order constraints, Vandekerckhove, Sandra, and Daelemans (2013) showed that the behavioural pattern of the patients with impaired adjective order knowledge described by Kemmerer et al.

(2009) could be explained as resulting from *overeager abstraction*, an impairment of the online abstraction process in which participants put too much weight on relatively distant distributional-semantic neighbours of the input words. However, their reanalysis of Kemmerer et al.'s (2009) data did not reveal any effect of online abstraction on the unimpaired participants' adjective order preferences. This full dissociation between the selectively impaired patients and the unimpaired participants suggests a qualitative difference between the generalization mechanisms of the two groups: unimpaired language users might normally refer to the persistent summary representations of abstract positional categories, and not apply online extrapolation and concomitant abstraction from lower-order representations. This might make them immune to effects of neighbour interference. The impaired participants' condition might have caused them instead to lose access to the positional categories, forcing them to rely on an alternative, online-abstraction mechanism, and making them overly sensitive to interfering neighbours.

The case for online abstraction as a general co-determinant of order sensitivity in the processing of prenominal adjective sequences would be stronger if one could provide evidence that effects tied to the process of online abstraction also surface in the adjective order preferences of unimpaired language users. In this study, we investigated whether such effects can be found for the order preferences of Dutch adjective pairs. More specifically, we set out to test the following two hypotheses:

(1) Online abstraction over adjective exemplars explains subregularities in participants' adjective order preferences, i.e. order preference variation that cannot be accounted for by either phrase-specific ordering preferences or the linear precedence between abstract positional categories.

(2) Online abstraction over adjective exemplars fully accounts for the adjective order regularity that is explained by the linear precedence between abstract positional categories.

The first hypothesis was tested using a 2AFC task in which participants had to choose between noun phrases containing differently ordered prenominal adjective pairs. The (stronger) second hypothesis was tested in a reanalysis of the 2AFC task. If both hypotheses are confirmed, online abstraction provides a single-route explanation of adjective order preferences, accounting for both the order preference variation that can be captured by the linear precedence relationship between the positional categories, and the variation in adjective order preferences that escapes those categorical precedence constraints. In the next two sections, we introduce our implementations of a category-based and an online abstraction approach to adjective order preferences, before discussing the model-based stimuli selection and the experiment itself.

A category-based model of prenominal adjective order: category precedence

According to category-based accounts of prenominal adjective order, the linear precedence relation between the respective positional categories of two adjectives predicts their preferred order. We will call this category-based order preference *category precedence.* In order to operationalize this variable, we translated a representative, comprehensive, and sufficiently specific linguistic description of prenominal adjective order, namely that of Bache and Davidsen-Nielsen (1997), into a categorization model that was trained on a large text corpus. This model assigned the adjectives to the appropriate categories using the category-associated properties provided by Bache and Davidsen-Nielsen (1997) as predictive features.

Adjective categories

Bache and Davidsen-Nielsen's (1997) description of the constraints on the order of prenominal adjectives in English is one of the most comprehensive of all existing category-based accounts (also see Bache, 1978). Their taxonomy is shown in Table 1.[1] At the highest level of premodifier organization, Bache and Davidsen-Nielsen (1997) distinguished three *modification zones*: *Specification* (MOD1), *Description* (MOD2), and *Classification* (MOD3). Descriptive adjectives are prototypical modifiers: they ascribe an attribute to the noun referent, e.g. *a long dress*, *a perilous endeavour*. Specifying adjectives are closer in function to the *determination* function of determiners, i.e. resolving the reference of the noun, e.g. *the other door*, *the largest dinosaurs*. Classifying adjectives are closest to the categorization function of nouns, e.g. *an industrial complex*, *a nervous disease*.

Within the different modification zones, there is a second level of more fine-grained distinctions. Table 1 lists 18 "traditional 'order classes'" (Bache, 1978, p. 32). These adjective-inherent order classes commonly reappear in different category-based accounts of prenominal adjective order (see e.g. Dixon, 1982; Hetzron, 1978; Scott, 2002). An adjective can occur in different zones depending on its modificational subfunction with respect to a specific noun in a given context, but if the adjective

[1] Bache and Davidsen-Nielsen's (1997) ordering scheme contained lexical categories that not all linguistic paradigms might regard as being subject to adjective ordering restrictions (i.e. quantifiers and nouns). According to such paradigms, changing the position of those categories relative to "true" modifiers invariably results in an ungrammatical order. The experiment reported in this paper does not address the question whether adjective order theories can also explain the order of quantifiers and prenominal nouns, as the items in our stimulus set did not contain words belonging to those categories (also see Footnote 2).

Table 1. Linear order of the premodifier modification zones and their subcategories, according to Bache and Davidsen-Nielsen (1997)

MOD1: SPECIFICATION	ORDINAL NUMBERS	*first*, *seventh*, *next*, *final*
	CARDINAL NUMBERS	*two*, *few*, *many*, *countless*
	COMPARED FORMS	*older*, *better-known*
	OTHER	*only*, *own*, *same*, *former*, *major*
MOD2: DESCRIPTION	EVALUATIVE	*good*, *nice*, *pretty*, *horrible*
	SIZE	*big*, *large*, *huge*, *vast*, *small*
	LENGTH	*long*, *short*
	HEIGHT	*tall*, *high*, *low*
	OTHER	*rough*, *hot*, *soft*, *heavy*, *young*
	DEVERBAL	*undulating*, *predictable*
	DENOMINAL	*hilly*, *wishful*, *dusky*, *ghoulish*
MOD3: CLASSIFICATION	DEVERBAL	*leading*, *internalized*
	COLOUR	*green*, *red*, *yellow*, *black*
	NATIONALITY	*English*, *Chinese*, *Scandinavian*
	LOCALITY/TIME	*local*, *Atlanta*, *daily*, *annual*
	DENOMINAL	*industrial*, *nuclear*, *woollen*
	NOMINAL	*metal*, *silk*, *foreign policy*
	NON-INHERENT	*wild* (*bird*), *secret* (*service*)

Note. EVALUATIVE corresponds to Bache and Davidsen-Nielsen's (1997) +EMOTIONAL.

belongs to one of the order classes, it will most frequently occur in the modification zone associated with that class, and in the within-zone position given in Table 1.

Following the taxonomy of Bache and Davidsen-Nielsen (1997) (see Table 1), adjectives were categorized at the level of the three modification zones, and separately assigned to one of the appropriate order classes. The order preference of an adjective pair was then determined by the linear precedence relationship between the Mod-zone–order-class combinations of the two adjectives.[2]

Adjective features

Adjectives were represented by 10 features. Features 1 to 8 were used for the classification of the adjectives into Mod-zones. Feature 9 was used to classify adjectives within MOD2. Feature 10 represented the adjective-inherent order classes. The first five features operationalized four properties that Bache and Davidsen-Nielsen (1997) listed as characterizing central adjectives. Feature 6 implemented the procedure given by Bache (1978) to establish the Mod-zone of prenominal adjectives on the basis of their position vis-à-vis unambiguous MOD2 adjectives. Feature 7 captured the general association of an adjective's preferred position in adjective pairs with the Mod-zones. Feature 8 was adapted from Warren (1984) and was meant to identify specifying

---

[2] The taxonomy and corresponding precedence hierarchy of Bache and Davidsen-Nielsen (1997) is applicable to Dutch with only one minor change: because bare nouns do not function as premodifiers in Dutch, there is no need for a MOD3: NOMINAL category. For a Dutch noun to function as a premodifier, it either needs to be derived into an adjectival modifier by appending the denominalization suffix *-en* (e.g. *het standvastige tin*nen *soldaatje* [the steadfast tin soldier]) or merged with the head noun into an orthographic compound (e.g. *cotton industry* translates to *katoenindustrie*).

adjectives in particular. The features are detailed below:

(1) *Comparison*: the number of (morphologically marked) comparative and (morphologically and lexically marked) superlative forms of the adjective. Together with the intensification feature below, the comparison feature captures the gradability of the adjective.

(2) *Intensification*: the number of times that the adjective is preceded by the intensifiers *zeer* [very] or *erg* [very, awfully].

(3) *Predication*: the number of times that the uninflected adjective is preceded by an active form of the copula verbs *zijn* [be] or *lijken* [seem] and followed by a period, question mark, or exclamation mark.

(4) *Coordination*: the number of times the inflected adjective forms a comma-separated pair with another (preceding or following) adjective in front of a noun, e.g. *aantrekkelijke, jonge vrouw* [attractive, young woman] or *jonge, aantrekkelijke vrouw* [young, attractive woman].

(5) *Contrastive pair*: a binary feature indicating whether the adjective or any of its synonyms has antonyms.

(6) *MOD2 insertion*: of all the instances in which an inflected ambiguous adjective forms an unbroken pair with any of the adjectives from a set of 54 inflected inherent MOD2 adjectives, the proportion of times the ambiguous adjective precedes the MOD2 adjective. This feature implements the Mod-zone identification strategy given by Bache (1978). If an adjective tends to precede inherent MOD2 adjectives, it is more strongly associated with MOD1. If it tends to follow inherent MOD2 adjectives, it is more strongly associated with MOD3. If an adjective has neither tendency, it is most strongly associated with MOD2.

(7) *Preferred position*: the proportion of times the inflected adjective occurs first when forming an unbroken pair with another adjective immediately preceding a noun.

(8) *Indefinite article*: the proportion of times the inflected adjective is immediately preceded by the indefinite article *een* [a/an]. This feature implements Warren's (1984) suggestion that phrases with specifying adjectives tend to be definite, e.g. *a tall woman* vs. *a tallest woman*.

(9) *Subjectivity*: A binary feature indicating whether an adjective is subjective or not. This feature was not used for the classification into Mod-zones, but to distinguish between +EVALUATIVE and −EVALUATIVE adjectives within MOD2.

(10) *Order class*: We followed Bache and Davidsen-Nielsen (1997) in treating the order classes at the level below the Mod-zones as adjective-inherent in a deterministic fashion. Separately from the classification into the three Mod-zones, adjectives were assigned to one of the order classes listed in Table 1.

Adjectives were manually tagged for the binary contrastive pair feature, using the online version of the *Van Dale Groot woordenboek van de Nederlandse taal* (den Boon & Geeraerts, 2005). As for the subjectivity feature, an adjective was considered to be subjective when it was listed as a "strongsubj" item in the *MPQA subjectivity* lexicon[3] (Wilson, Wiebe, & Hoffmann, 2005). Adjectives were also manually tagged for the order class feature; if the adjective did not clearly belong to one of the established order classes, it was assigned to the OTHER class. The values for the other, numeric features were automatically estimated from the 1999–2002 section of the *Twente News Corpus* (TwNC) (Ordelman, de Jong, van Hessen, & Hondorp, 2007). This corpus

[3] Available from http://mpqa.cs.pitt.edu/lexicons/subj_lexicon.

contains over 300 million words from Dutch periodicals and teleprompter files. Except for the comparison feature, the relevant input from the corpus was first tagged with parts of speech (POS) to minimize miscounts due to lexical ambiguities. For that purpose, we used the POS-tagging functionality of *Frog*[4] (van den Bosch, Busser, Daelemans, & Canisius, 2007), a collection of memory-based natural language processing modules for Dutch. To make the feature values comparable between different adjectives, the raw frequencies were divided by the overall corpus frequency of the respective inflected adjective (this applied to all numeric features except MOD2 insertion and preferred position). The comparison, intensification, predication, coordination, and indefinite article features were positively skewed, so they were brought closer to normality using Box–Cox transformations (Box & Cox, 1964), as implemented in the *R* package *geoR* (Diggle & Ribeiro Jr, 2007; Ribeiro Jr & Diggle, 2001). Before transforming the features, add-one smoothing was applied to avoid zero counts.

Classification algorithm

Bache and Davidsen-Nielsen (1997) stressed that the descriptive, specifying, and classifying zones are not adjective subclasses but functional categories. Yet, they also acknowledged the requirement to "regard adjectives as *inherent* MOD1, MOD2 or MOD3 adjectives, according to their typical usage" (p. 465). Although Bache and Davidsen-Nielsen (1997) did not resolve this opposition, the apparent tension between a functionally determined and an adjective-inherent treatment of the Mod-zones disappears when treating the categorization of adjectives into those zones

---

[4] Available from http://ilk.uvt.nl/frog/.

probabilistically instead of deterministically, i.e. by implementing the characteristics listed by Bache and Davidsen-Nielsen (1997) as probabilistic constraints instead of absolute rules. In a probabilistic approach, if the context is neutral, the opposition between categories either being inherent or functionally determined becomes irrelevant. Instead of having the classifier assign each adjective to a specific category, it outputs a probability distribution over the categories to provide a probabilistic order preference. An adjective whose probability mass is skewed towards one specific Mod-zone then simply has a stricter order preference than an adjective whose probability mass is evenly distributed over Mod-zones.

To categorize adjectives as MOD1, MOD2, or MOD3, we employed a machine learning approach that determined the relative predictive value of the category criteria based on the first eight features listed in the previous section. Because Bache and Davidsen-Nielsen (1997) provided a number of example adjectives for each of the order classes inherently associated with one of the three Mod-zones, it was possible to use a supervised classification algorithm. More specifically, we used a Support Vector Machine (SVM) classifier (Cortes & Vapnik, 1995) with a Gaussian radial basis function kernel, as implemented in the *R* package *e1071* (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2012). The SVM was trained on a manually categorized set of adjectives that were either listed by Bache and Davidsen-Nielsen (1997) as inherently being associated with one of the three Mod-zones, or belonged to an order class with that same association (e.g. SIZE and COLOUR are inherently MOD2 and MOD3 classes, respectively).

For the determination of the Mod-zone–order-class combinations, the interpretation of the adjectives' order classes (feature 10 above) shifted depending on the Mod-zone being considered. If the order class was not listed under the specific Mod-

zone in Bache's scheme, the adjective was classified as OTHER for MOD1 and MOD2, and as NON-INHERENT for MOD3. For example, in combination with MOD2, a SIZE adjective was taken to be MOD2: SIZE. However, under MOD1 and MOD3, it was taken to be MOD1: OTHER and MOD3: NON-INHERENT, respectively. Within MOD2, subjective adjectives were always considered MOD2: EVALUATIVE, regardless of their order class. So being a subjective denominal, *redelijke* [reasonable] was considered OTHER under MOD1, EVALUATIVE under MOD2, and DENOMINAL under MOD3.

Category precedence

To determine the category precedence of an adjective bigram $adj_1 adj_2$, we took the Cartesian product of both adjectives' discrete probability distributions over Mod-zones, and multiplied the two probabilities of each pair in the product set. Each of those compound probabilities was multiplied by a factor indicating the order precedence between the two Mod-zone–order-class combinations in question. This factor equalled 1 if the Mod-zone–order-class combination of adj_1 preceded the Mod-zone–order-class combination of adj_2 on the linear precedence scale, 0.5 if both category combinations were the same, and 0 if the Mod-zone–order-class combination of adj_1 followed the Mod-zone–order-class combination of adj_2 on the precedence scale. The category precedence of an adjective bigram was given by the sum of those compound probabilities. Accordingly, category precedence ranged from 0 to 1, with 1 being a maximal preference for the given order of an adjective pair, and 0 being a maximal preference for the inverse order.

An online abstraction account of adjective order preferences: neighbourhood support

Whereas a category-based account explains the order preference of an adjective pair in terms of category precedence, online abstraction predicts that the preference for a specific adjective order is dependent on the order's *neighbourhood support*. The neighbourhood support for an adjective order $adj_1 adj_2$ was operationalized as the number of word types in the intersection between the set of four words most similar to adj_2 (adj_2's neighbourhood) and the set of words that immediately follow any of the four words most similar to adj_1 in the TwNC. The similarity between two adjectives was calculated as the difference between the distributions of the words following those adjectives in the training corpus. Because the word most similar to a given adjective is the adjective itself, adj_1 and adj_2 were always contained in their own neighbourhoods.

Because the neighbourhood support for a specific adjective order was the number of items in a subset of adj_2's four nearest neighbours, it could range from zero to four. A neighbourhood support value of four meant that all four neighbours in the neighbourhood of adj_2 were preceded by at least one of the nearest neighbours of adj_1 in the training corpus. A neighbourhood support value of zero meant that none of the neighbours of adj_2 were among the words that followed the nearest neighbours of adj_1.

The size of adj_2's neighbourhood was fixed at four to allow for variation in the size of the intersection between that neighbourhood and the set of words following adj_1's nearest neighbours, while still limiting the neighbourhood to words that were reasonably similar to adj_2. The fact that the neighbourhood of adj_1 also had a size of four words was only a matter of convenience, as setting both neighbourhood sizes to the same value simplified the implementation of the exemplar model.

To create the adjectives' neighbourhoods, the IB1 memory-based learning algorithm in the TiMBL software package[5] (Daelemans, Zavrel, van der Sloot, & van den Bosch, 2010) was trained on an exemplar database that consisted of adjective-initiated word surface form bigrams from TwNC.[6] The version of TwNC that was used for the composition of the exemplar memory had already been POS-tagged using a memory-based tagger (Daelemans, Zavrel, van den Bosch, & van der Sloot, 2003) trained on a version of the Eindhoven Corpus (Uit den Boogaart, 1975) that was itself tagged with the WOTAN tag set (Berghmans, 1994). To keep computational demands low, both the adjective and the following word, or their lowercased equivalents, had to be listed in the CELEX (Baayen, Piepenbrock, & Gulikers, 1995) word forms database to be included in the data set.[7] While the first words of the bigrams (w_i) were restricted to adjectives, the second words (w_{i+1}) could be any part of speech. This exemplar database contained 13,528,982 bigram tokens, and 2,335,046 bigram types. The w_i and w_{i+1} type counts were 21,682 and 108,735, respectively.

Experiment: Two-alternative forced choice

With this experiment, we tested the hypothesis that there is variation in adjective order preferences that is not accounted for by the linear precedence constraints between the adjectives' positional categories, but can be explained by the order's *neighbourhood*

---

[5] Available from https://ilk.uvt.nl/timbl/.

[6] The IB1 algorithm used the following settings: a neighbourhood size (k) of 4, tie resolution by random class selection from the exemplar neighbourhood, no feature weighting, and Modified Value Difference Metric (MVDM) (Cost & Salzberg, 1993) (with w_{i+1} as the exemplar class) for distance metric. Other parameters were kept at their default values.

[7] We used the *Python*™ package *Leanlex* (Keuleers, 2007) to access CELEX.

support, i.e. the support for the respective orders among the nearest neighbours of the adjectives in the test items. Like Kemmerer et al. (2009), we used a 2AFC task requiring participants to choose between two noun phrases that differed only in the order of the two adjectives that preceded the noun. However, in this case, the adjective bigrams contained in the test items were explicitly selected to vary in the extent to which their order was supported by their nearest neighbours in an exemplar memory of adjective-initiated bigrams. We hypothesized that there would still be a positive effect of neighbourhood support on the probability that participants selected the target order in a model that already included the effect of *category precedence*, i.e. the order preference predicted by the linear precedence relationship between the positional categories of the two adjectives.

Materials and methods

Materials

We created 75 Dutch noun phrase pairs of the form *det* adj_a adj_b *noun* vs. *det* adj_b adj_a *noun*. The adjective bigrams were chosen in such a way that one order, the *control order*, always had a neighbourhood support of 4 out of 4, whereas the neighbourhood support for the other order, the *target order*, ranged from 0 to 4. The stimulus noun phrases and associated variable values are provided as online supplemental material.

Taking the set of w_i adjectives and the subset of w_{i+1} words tagged as nouns from the *adjective noun* bigrams in the exemplar database extracted from the TwNC (see above), we first generated a set of (adj_a, adj_b, *noun*) combinations (with $adj_a \neq adj_b$) from which to pick the actual test items. Those triads satisfied the following constraints:

- The frequency per million of the adjective lemmas was at least 30 according to the Dutch lemma database of CELEX. 424 adjectives from the TwNC exemplar database satisfied this constraint.
- The degree of comparison of the adjectives was the positive degree; comparative and superlative adjectives were excluded.
- Both adjectives ended with the -*e* suffix. Apart from reducing surface form variation, this constraint minimized the possibility that participants interpreted the adjective pairs as adverb–adjective combinations (as the base forms of adverbs and adjectives are not formally different in Dutch).
- The length of the adjectives did not differ by more than two letters. There is evidence for English that shorter adjectives tend to precede longer adjectives (Wulff, 2003).
- Both adj_a adj_b and adj_b adj_a were unattested in the TwNC, which minimized the chance that participants retrieved the adjective bigrams as lexical chunks.
- Both adj_a *noun* and adj_b *noun* were attested in the TwNC. That way, a proper semantic fit between each adjective and the noun was guaranteed. Additionally, the reverse conditional bigram probabilities $\mathrm{P}(adj_a|noun)$ and $\mathrm{P}(adj_b|noun)$ were equal, to control for the adjectives' strength of association with their head nouns. It has been shown that $\mathrm{P}(adjective|noun)$ is higher for adjectives that occur closest to the head noun (Wulff, 2003).

Of the generated (adj_a, adj_b, *noun*) triads, 27,238 satisfied all of the above constraints. We ran both adjective orders of those triads through the trained exemplar model described above to obtain their neighbourhood support values. We kept those combinations of which one adjective order (the control order) had a neighbourhood

support of 4 out of 4 and whose neighbourhood support for the other order (the target order) was contained in the [0, 4] interval. The high support for the control order maximized the semantic felicitousness of the adjective combinations (resulting in the exclusion of contradictory pairs like *black white*), despite the constraint that they could not occur as bigrams in the reference corpus. From the thus created set of (adj_a, adj_b, *noun*) combinations, we selected 75 test items, i.e. 15 items for each of the 5 target order support levels. The 75 selected combinations were those whose adjective surface forms differed the least in both word length (measured in letters) and SUBTLEX-NL (Keuleers, Brysbaert, & New, 2010) log frequency count. Controlling for the frequency difference was motivated by the finding that there is a tendency for more frequent prenominal adjectives to precede less frequent ones (Wulff, 2003). Consequently, the letter length distributions did not differ between adj_a and adj_b (both $Mdn = 7$) according to a Wilcoxon rank sum test, $W = 3080$, $p = .309$ two-tailed, nor did their log frequencies ($Mdn = 2.65$ and $Mdn = 2.69$, respectively), $W = 2629.5$, $p = .493$ two-tailed.

Although we aimed at minimizing the number of repeated adjectives, some recurrence of adjectives over different test items turned out to be unavoidable. The 150 adjective slots were filled by 88 adjectives: 50 adjectives occurred once, 22 adjectives twice, 10 adjectives occurred three times, 4 adjectives four times, and 2 adjectives five times.

Upon selection of the 75 experimental items on the basis of their neighbourhood support, they were given a category precedence value by the categorical adjective order model described above. The Mod-zone SVM classifier was trained on a set of 111 adjectives that had been manually annotated with one of the three Mod-zones. Of those adjectives, 57 belonged to the set of experimental items, and an additional 54 were

taken from Bache and Davidsen-Nielsen's (1997) examples of adjectives with an inherent Mod-zone association. The γ and C classifier parameters were determined using a grid search over the ranges $[2^{-15}, 2^{-4}]$ and $[2^{-5}, 2^{15}]$, respectively. The resulting parameter settings were a γ of 1.95×10^{-3} and a C value of 256. We applied the trained classifier to all adjectives in our experimental stimuli, including the ones that we had tagged manually and that were part of the classifier's training set, to obtain three Mod-zone probabilities for each adjective.

The stimuli were not explicitly selected to vary in their category precedence independently of their target order neighbourhood support. Consequently, category precedence did correlate with neighbourhood support, $\tau = 0.28$, $p < .001$. Nevertheless, there was still considerable residual variation in the items' neighbourhood support, as shown in Figure 1. The category precedence of the 75 stimuli ranged from 0.01 to 0.98, with a median of 0.18 and a mean of 0.31. The difference between mean and median indicates that the variable had a positive skew.

To present the test items, determiners were added in front of the *adjective adjective noun* sequences. Items were counterbalanced across two lists such that the target order appeared in the first position in list one and in second position in list two for half of the items, and the other way around for the other half. The order of the items was randomized between participants.

Procedure

The task was administered over the Internet, using the survey tool *LimeSurvey* (LimeSurvey Project Team & Schmitz, 2011). Participants were invited by email and could carry out the task at their own convenience. The noun phrase pairs were presented in list form, with 15 pairs on each page, and the two orders of each pair shown one

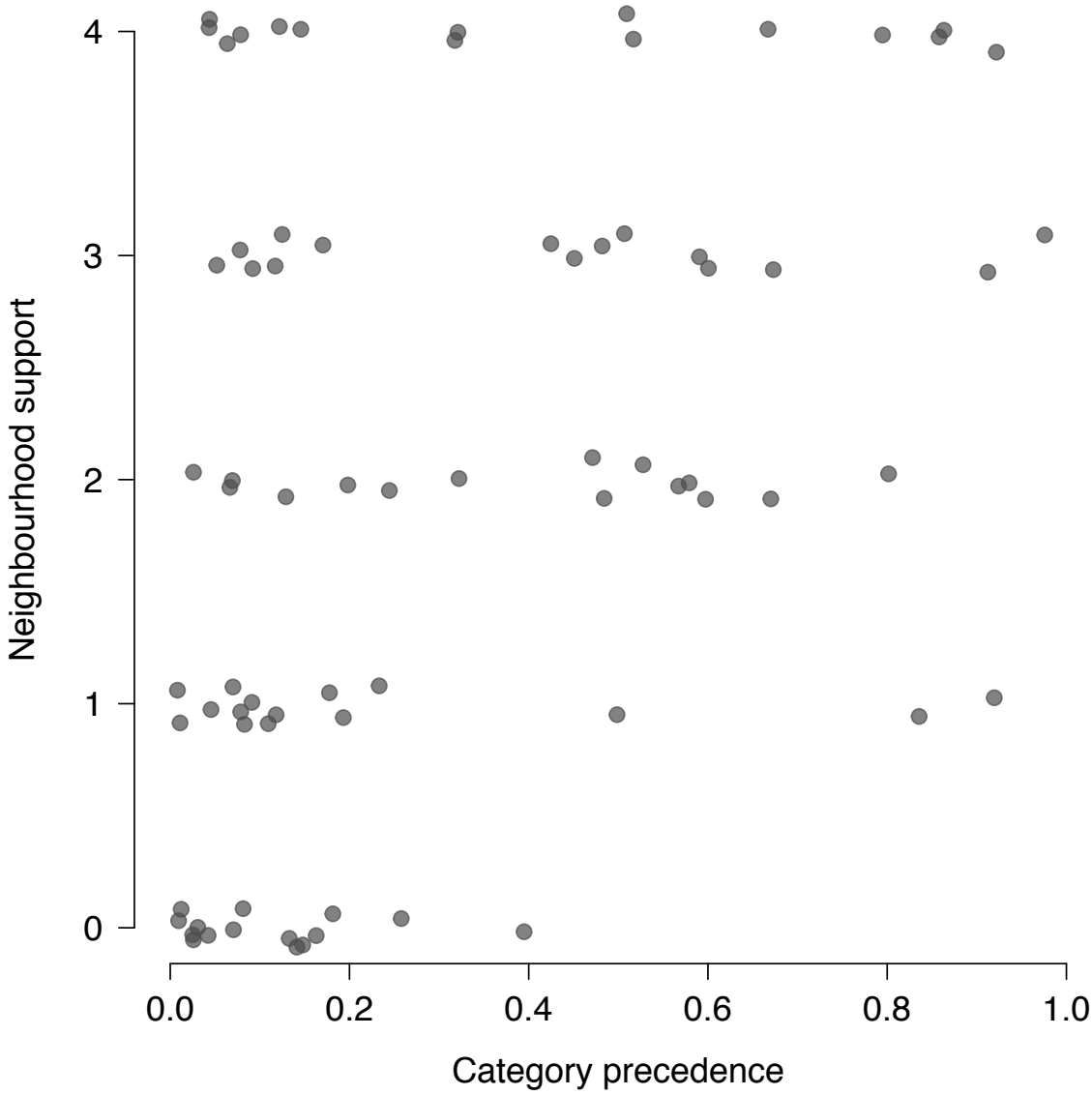

Figure 1. Distribution of category precedence (x-axis) versus neighbourhood support (y-axis) for the 75 items in the stimulus set.

below the other. To the left of each noun phrase was a radio button. The participants' task was to indicate which of the two orders they preferred for each pair by clicking on the appropriate radio button. The instructions encouraged the participants to read both phrases carefully before making a decision, and to rely on their intuition. There were no practice items.

Participants

35 students (24 female and 11 male) from the *Linguistics and Literature* program at the University of Antwerp participated in this experiment. The mean age of the participants was 20.11 (*SD* = 2.51). All participants reported to be native speakers of Dutch.

Analysis method

The results of the forced choice task were analysed with mixed-effects logistic regression, using the *R* package *lme4* (Bates, Mächler, Bolker, & Walker, 2014). The dependent variable was *selected adjective order*, with the levels *target* and *control*. The fixed-effects independent variables of interest were neighbourhood support, category precedence, and their interaction. Because category precedence had a positive skew, the variable was Box–Cox transformed before it was entered into the model. All variables were centred. To obtain more accurate estimates, we also included the binary variable *target order position* as a fixed-effects predictor. This variable indicated whether the target order was presented as the first or second item of each stimulus pair. Random-effects variables that were justified for inclusion in the model by the experimental design were random intercepts for subject, item, adj_1 (the first adjective in the target order), adj_2, by-subject random slopes for all fixed-effects variables, and by-item, by-adj_1, and by-adj_2 random slopes for target order position. Random effects for adj_1 and adj_2 were motivated by the limited repetition of adjectives between items. To determine the random effects structure, we first considered a model that included all the random-effects variables justified by the experimental design, and fell back on a forward best-path algorithm to obtain the maximum random effects structure supported by the data if the full model did not converge (Barr, Levy, Scheepers, & Tily, 2013). This algorithm started from a model that included all fixed-effects independent variables and by-subject

random intercepts, and used an α of .20 for both inclusion and exclusion of the random-effects variables.

Results and discussion

There were 2625 observations, i.e. 35 participants × 75 items. Because the interface did not allow participants to skip items, no observations were missing. A model with the full random-effects structure justified by the experimental design failed to converge. The forward best-path algorithm resulted in a model with random intercepts for subject and adj_2, random by-adj_1 slopes for category precedence, and random by-subject slopes for target order position.

A visual summary of the experimental results is given in Figure 2. The results of the regression analysis are presented in Table 2. The estimated overall probability of participants selecting the target order was 0.23. Target order position had a small positive effect on the probability that the target order was selected. With all other independent variables held constant, when the target order was presented as the first of the two orders in a stimulus pair, the odds of participants selecting the target order was 30% higher than when the target order was presented as the second item. Both category precedence and neighbourhood support had a positive effect on selected adjective order, but did not interact. Because of the Box–Cox transformation on category precedence, the effect of this variable cannot easily be interpreted on the basis of the estimated coefficients. Figure 3 therefore visualizes the estimated effects of category precedence on selected adjective order, after transforming the latter variable back to its original scale. The effect of neighbourhood support is also shown in this figure. With target order position and category precedence held constant, the odds of participants selecting

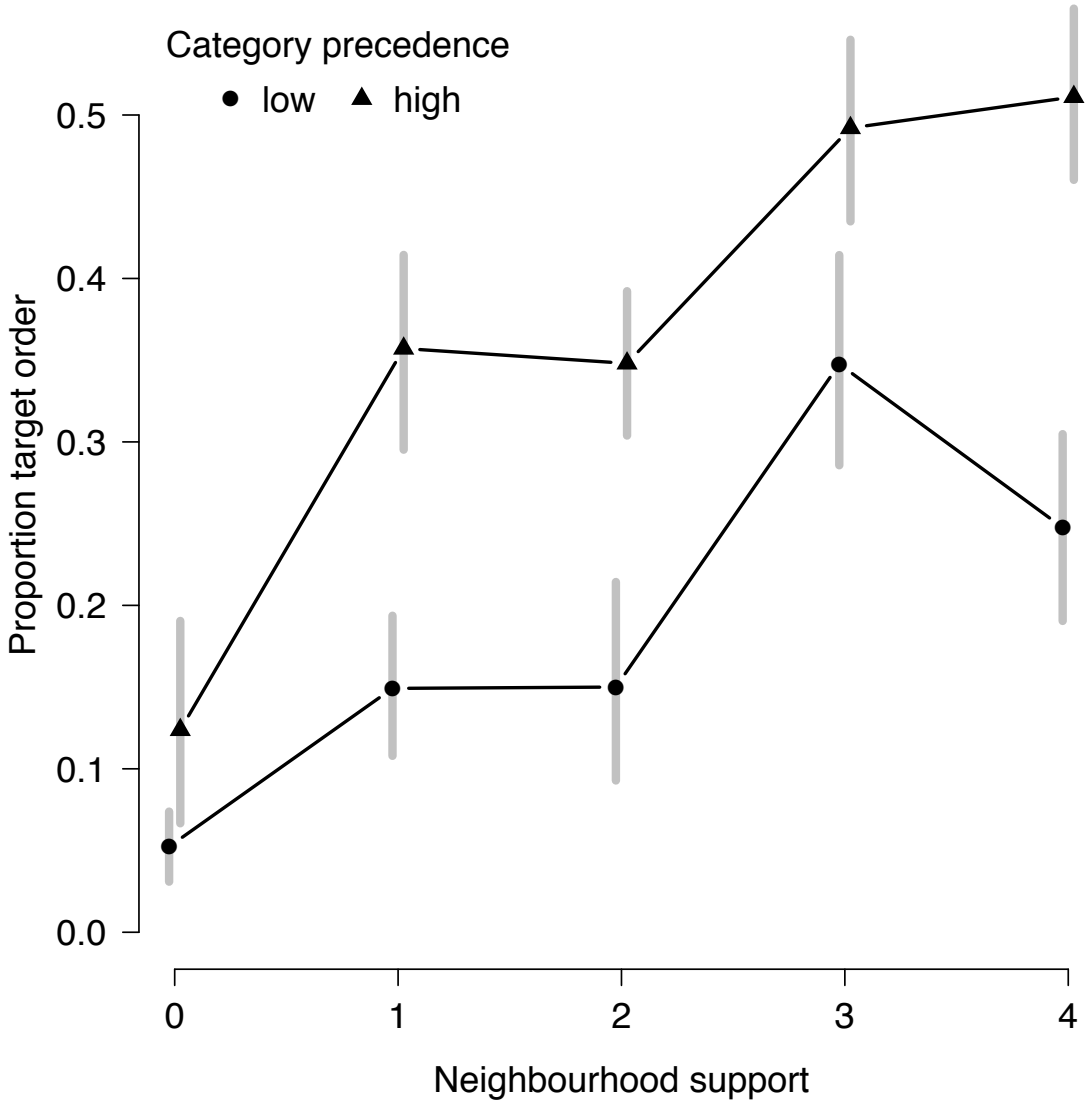

Figure 2. Proportion of trials in which participants selected the target order for the five levels of neighbourhood support and by median-split category precedence (*median* = 0.18), with non-parametric bootstrap 95% confidence intervals. Note that the confidence intervals ignore the participants × items structure.

the target order increases with 60% for every one-neighbour increase of the target order's neighbourhood support.

Because the stimuli were created such that they contained only low-frequent adjective pairs, it is unlikely that the order preferences of the participants stemmed from their memory of the specific adjective bigrams in the stimuli. Instead, the participants

Table 2. Estimated fixed effects of a mixed-effects logistic regression model predicting selected adjective order from target order position, category precedence, and target order neighbourhood support

Parameter	B	e^B	$SE(B)$	z	χ^2	$p(>\chi^2)$
Constant	−1.23	0.29	0.16	−7.64	42.65	< .001
Target order position	0.26	1.30	0.12	2.11	4.15	.035
Category precedence	1.16	3.20	0.23	5.14	21.49	< .001
Neighbourhood support	0.47	1.60	0.10	4.87	21.24	< .001
Category precedence × Neighbourhood support	−0.14	0.86	0.15	−0.98	0.97	.327

Note. All independent variables in the model are mean-centred.

relied on generalized adjective order constraints by abstracting away from the specific adjectives in the stimulus phrases. Category-based accounts capture the abstractions that enable generalization of adjective order preferences in a set of positional categories, and predict that adjective order preferences should correlate with the linear precedence relationship between those categories. Indeed, category precedence explained an important part of the variation in the participants' order choices. However, on top of the category precedence effect, we found that the amount of support for a specific adjective order in the neighbourhoods of the adjectives also captured selected order variation. This confirms the hypothesis that online abstraction over adjective exemplars accounts for variation in adjective order preferences that cannot be explained by the linear precedence between abstract positional categories.

Apart from the order preference variation that was uniquely accounted for by neighbourhood support, the variable also competed for explained variance with category

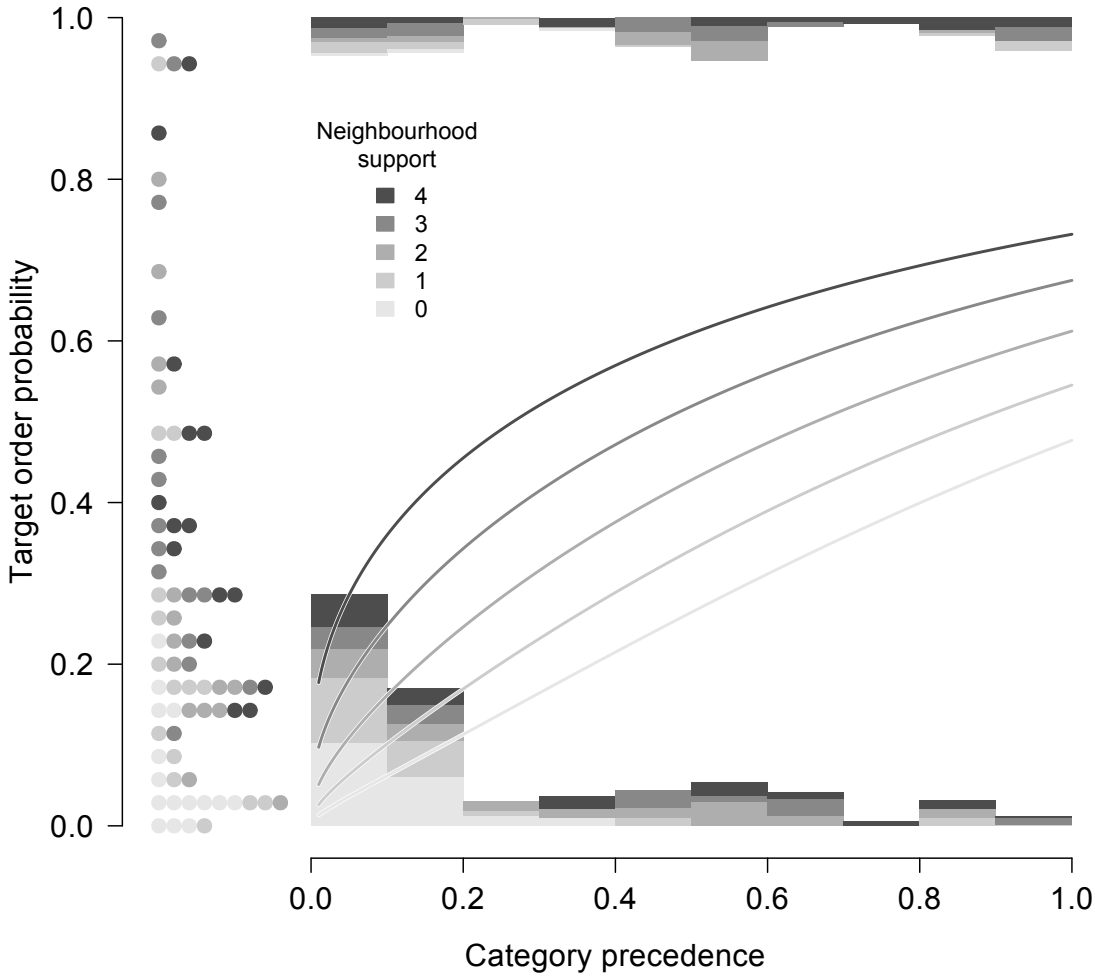

Figure 3. Partial effects of neighbourhood support and category precedence on selected adjective order, at the median target order position of 0.5. The stacked histograms show the distribution of responses over both independent variables (top = target order, bottom = control order). The stacked strip chart along the y-axis shows the by-item target order proportions.

precedence, and did not eliminate the need for the latter as an explanatory variable. This suggests that the effects of the exemplar-based and the category-based variables were grounded in two qualitatively different processing mechanisms. Based on these findings, the most parsimonious model for explaining the generalization of adjective order preferences is one in which the linear precedence relationship between established

categories and the support of online abstractions over adjective exemplars play additive roles.

Reanalysis of the experiment

Before accepting the dual-mechanism explanation, however, one needs to consider the possibility that neighbourhood support was not operationalized in the most optimal way. For instance, there was no empirical or theoretical reason to cut off neighbourhood size exactly at 4 for all items. Additionally, there is more information available in the neighbourhood of the adjectives than was incorporated into the original operationalization.

We therefore reanalyzed the 2AFC task results using a number of more fine-grained exemplar-based variables. The goal of this reanalysis was to investigate the hypothesis that online abstraction not only adds to the explained variation in adjective order preferences on top of the variation captured by the linear precedence relationship between abstract positional categories, but that online abstraction over concrete exemplars can fully explain away the effect of category precedence.

Materials and methods

Alternative operationalizations of online abstraction

In this section, we introduce four variables that incorporate more detailed information about the stimulus adjectives' neighbourhoods. The variables are grouped two-by-two theoretically. The first two variables, grouped under *nearest bigram support*, capture the distance-weighted conditional bigram probabilities of the adjectives' nearest neighbour bigrams, for the target and control order, respectively. The last two variables, grouped under *positional precedence*, capture the respective probabilities of the two adjectives

occurring in first position when paired, weighted by function of the other adjective's neighbours. To estimate the values of all four variables, the same exemplar memory and similarity measure were used as for the original operationalization of neighbourhood support.

Nearest bigram support. The original operationalization of neighbourhood support captured how many of the four nearest neighbours of adj_2 followed any of the four nearest neighbours of adj_1. Apart from the fact that there was no empirical or theoretical reason to select the specific neighbourhood size of four, the operationalization did not take into account the exact probability of those adj_2 neighbours following the specific adj_1 neighbours, nor how distant the adjective neighbours actually were from the adjectives in the stimulus phrases. Given that the effect of neighbourhood support was established in the 2AFC experiment, a logical next step is to hypothesize that (a) a neighbour bigram with a high conditional bigram probability will provide stronger support for a specific adjective order than a neighbour bigram with a low conditional probability, and (b) an order will be more strongly preferred if the nearest supporting neighbours are at a small distance from the focus adjectives than when they are at a larger distance. The more fine-grained differentiation between items that those variables provide might correlate with more fine-grained variation in the participants' responses. To operationalize the exemplar support for an adjective bigram adj_a adj_b in terms of nearest neighbour distance and the conditional bigram probability of neighbour bigrams, we considered two new variables:

- $P_\delta(adj_b|NN_{adj_a})$: the conditional bigram probability of adj_b, conditioned on the nearest neighbour of adj_a that precedes adj_b in the training corpus (NN_{adj_a}), and weighted by function of the distance between adj_a and NN_{adj_a}.
- $P_\delta(NN_{adj_b}|adj_a)$: the distance-weighted conditional bigram probability of adj_b's nearest neighbour that follows adj_a in the bigram exemplar memory (NN_{adj_b}), conditioned on adj_a.

To incorporate the role of neighbour distance into these nearest bigram probabilities, they were weighted by function of the distance to the nearest supporting neighbour of adj_a or adj_b, depending on the adjective that is smoothed over. We used *inverse-linear* (IL) distance weighting (Dudani, 1976). In this weighting scheme, the nearest neighbour receives a weight of 1, and the farthest a weight of 0. The weights of the neighbours in between those extremes are scaled linearly to the [1, 0] interval,

$$w_j = \begin{cases} \dfrac{d_k - d_j}{d_k - d_1} & \text{if } d_k \neq d_1 \\ 1 & \text{if } d_k = d_1 \end{cases}, \tag{1}$$

where d_j is the distance between the query and the *j*'th nearest neighbour, d_1 the distance of the nearest neighbour, and d_k the distance to the neighbour at the farthest *k*. Using scaled distance weights instead of absolute distance weights was necessary to match the effect of distance weighting between adjectives, as the absolute neighbour distances of the adjectives in the stimulus set were strongly correlated within adjectives.

The largest neighbourhood size required to find a non-zero probability for a $NN_{adj_a}adj_b$ or $adj_a NN_{adj_b}$ bigram was 703. Using smoothing over the first adjective of the bigram *buitenlandse gebruikelijke* [foreign usual]—in other words, estimating $P_\delta(gebruikelijke|NN_{buitenlandse})$—the nearest neighbour of *buitenlandse* that preceded

gebruikelijke in the exemplar database was its 703th neighbour, i.e. *laatste* [last]. For the estimation of the nearest bigram probabilities, we therefore weighted the probabilities according to a rescaling of the distances between the 2nd and the 703th nearest neighbour to the interval [1, 0] (the neighbour at $k = 1$ is always the adjective itself, with a distance of 0).

Because $\mathrm{P}_\delta(adj_b|NN_{adj_a})$ and $\mathrm{P}_\delta(NN_{adj_b}|adj_a)$ can be measured for both the target and the control order, in fact four variables were initially created to measure the exemplar support for the adjective pair with target order $adj_1\ adj_2$, namely:

- $\mathrm{P}_\delta(adj_2|NN_{adj_1})$
- $\mathrm{P}_\delta(NN_{adj_2}|adj_1)$
- $\mathrm{P}_\delta(adj_1|NN_{adj_2})$
- $\mathrm{P}_\delta(NN_{adj_1}|adj_2)$

Within the 75 test pairs of the 2AFC task, those variables are correlated, as shown in Table 3. To eliminate multicollinearity effects, we paired $\mathrm{P}_\delta(adj_2|NN_{adj_1})$ with $\mathrm{P}_\delta(NN_{adj_2}|adj_1)$ and $\mathrm{P}_\delta(adj_1|NN_{adj_2})$ with $\mathrm{P}_\delta(NN_{adj_1}|adj_2)$, orthogonalized the paired predictors using principal components analysis (PCA), and used the first component of each principal components object as explanatory variables in the regression analysis reported below. The original variables were Box–Cox transformed and converted to *z*-scores before they were entered into the PCA. The new variables resulting from the orthogonalizations are called *target* and *control nearest bigram support*.

Table 3. Pairwise correlations between the different distance-weighted conditional bigram probability variables

	$P_\delta(adj_1 \mid NN_{adj_2})$	$P_\delta(NN_{adj_2} \mid adj_1)$	$P_\delta(adj_2 \mid NN_{adj_1})$
$P_\delta(NN_{adj_2} \mid adj_1)$	−.03		
$P_\delta(adj_2 \mid NN_{adj_1})$	.07	.27***	
$P_\delta(NN_{adj_1} \mid adj_2)$	.26**	−.16*	−.16*

Note. The correlation strengths are given in Kendall's τ.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Distance-weighted positional precedence. The neighbourhood support and nearest bigram support variables do not directly predict adjective order preferences. A score for an adjective bigram with a particular order does not say much about the preferred order of the adjective pair unless it is compared to the score for the inverse order (pairings of adjectives that are unlikely to co-occur will have low scores for both orders). The most direct exemplar approach to estimate the preference for the target order $adj_1 adj_2$ is simply to estimate $P(adj_1 \prec adj_2 | adj_1, adj_2)$ on the basis of the co-occurrences of adj_1 and adj_2 in the training data. That approach cannot be applied to the adjective pairs that were used as test items; to have the participants abstract away from the specific adjectives in the stimuli, those pairs did not occur in the training data in either order. An exemplar approach explains this abstraction as the online matching of the input to its nearest, i.e. most similar, neighbours in memory. To determine the likelihood of a specific order given two adjectives, one can therefore refer to the neighbour adjectives of the adjectives in the stimulus, and estimate a similarity-smoothed probability. Instead of smoothing over both adjectives simultaneously and potentially attributing too much weight to one or the other adjective, we created two separate variables: adj_1 *precedence*,

$\mathrm{P}(adj_1 \prec N^k_{adj_2} | adj_1, N^k_{adj_2})$, and adj_2 *precedence,* $\mathrm{P}(adj_2 \prec N^k_{adj_1} | adj_2, N^k_{adj_1})$. The first variable captured the probability that adj_1 takes first position given that it is paired with any of the *k* nearest adjective neighbours of adj_2 ($N^k_{adj_2}$). The second variable captured the probability that adj_2 takes first position given that it is paired with any of the *k* nearest adjective neighbours of adj_1 ($N^k_{adj_1}$). Instead of setting *k* to an arbitrary value, or optimizing *k* on the set of test bigrams, again potentially overfitting the data, *k* was set to include all exemplars in the exemplar memory, and the specific $adj_1 N_{adj_2}$ and $adj_2 N_{adj_1}$ counts were weighted using the IL scheme (see Equation 1).

Analysis method

To assess whether online abstraction can account for all variation in selected adjective order that was previously explained by the effect of category precedence, we performed a two-step hierarchical mixed-effects logistic regression analysis. In the first step, a model was fitted that included the five exemplar-based variables as fixed effects, i.e. neighbourhood support, target nearest bigram support, control nearest bigram support, adj_1 precedence, and adj_2 precedence; and the target order position covariate (Model 1). The category precedence variable was added in the second step (Model 2). The random effects structure was the same for both models, and was determined by a best-path algorithm with an α of .20 that started from a model with random intercepts for subject and item, and all fixed-effects variables from Model 2. Random-effects variables that were considered for inclusion were random intercepts for subject, item, adj_1 and adj_2, random by-subject slopes for all Model 2 variables, and by-item, by-adj_1, and by-adj_2 random slopes for target order position. All independent variables were mean-centred.

Results and discussion

Table 4 gives the results of the hierarchical regression analysis. The summary of Model 1 shows that, of the four newly introduced exemplar-based variables, three had an effect on selected adjective order, all in the expected direction. Both target order nearest bigram support and adj_1 precedence had a positive effect on the odds of selecting the target order, i.e. higher values of those variables lead to a higher probability of the target order being selected. Conversely, adj_2 precedence had a negative effect on the odds of the target order being selected: the stronger the distance-weighted tendency of the target order's second adjective to be in first position, the lower the probability that participants selected the target order. The nearest bigram support for the control order did not have an effect on selected adjective order, because it was indirectly controlled for by keeping the control order's neighbourhood support constant at a value of four for the selection of the adjective pairs. The estimated odds ratio of neighbourhood support was brought back to 1 by the introduction of the exemplar-based variables. This means that the more fine-grained exemplar-based variables introduced in this section accounted for all variation in selected adjective order that was previously accounted for by neighbourhood support, effectively making that latter variable redundant.

The additional exemplar-based variables did not, however, account for all variation previously explained by category precedence, as shown in the Model 2 part of Table 4. They did strongly reduce the estimated effect size of category precedence, which went from an odds ratio of 3.20 in a model only containing neighbourhood support and category precedence (see Table 2) to an odds ratio of 1.55, but did not eliminate it. The estimated partial effects of Model 2 are shown in Figure 4.

Table 4. Hierarchical mixed-effects logistic regression analysis predicting selected adjective order from target order position, neighbourhood support, nearest bigram support for target and control order, adj_1 and adj_2 precedence, and category precedence

Parameter	B	e^B	$SE(B)$	z	χ^2	$p(>\chi^2)$
Model 1						
Constant	−1.33	0.26	0.13	−10.18	54.39	< .001
Target order position	0.26	1.30	0.13	2.10	4.15	.042
Neighbourhood support	0.03	1.03	0.10	0.27	0.07	.790
Target nearest bigram	0.66	1.93	0.13	5.00	24.15	< .001
Control nearest bigram	0.05	1.05	0.15	0.31	0.10	.753
adj_1 precedence	1.61	4.99	0.31	5.21	26.84	< .001
adj_2 precedence	−1.97	0.14	0.44	−4.43	18.47	< .001
Model 2						
Constant	−1.31	0.27	0.12	−11.15	61.85	< .001
Target order position	0.26	1.30	0.14	2.10	4.13	.042
Neighbourhood support	0.05	1.05	0.09	0.54	0.29	.592
Target nearest bigram	0.69	2.00	0.12	5.55	29.45	< .001
Control nearest bigram	0.07	1.07	0.15	0.45	0.20	.653
adj_1 precedence	1.23	3.43	0.31	3.97	16.44	< .001
adj_2 precedence	−1.53	0.22	0.44	−3.49	11.71	< .001
Category precedence	0.44	1.55	0.14	3.22	9.32	.002

Note. All independent variables in the models are mean-centred.

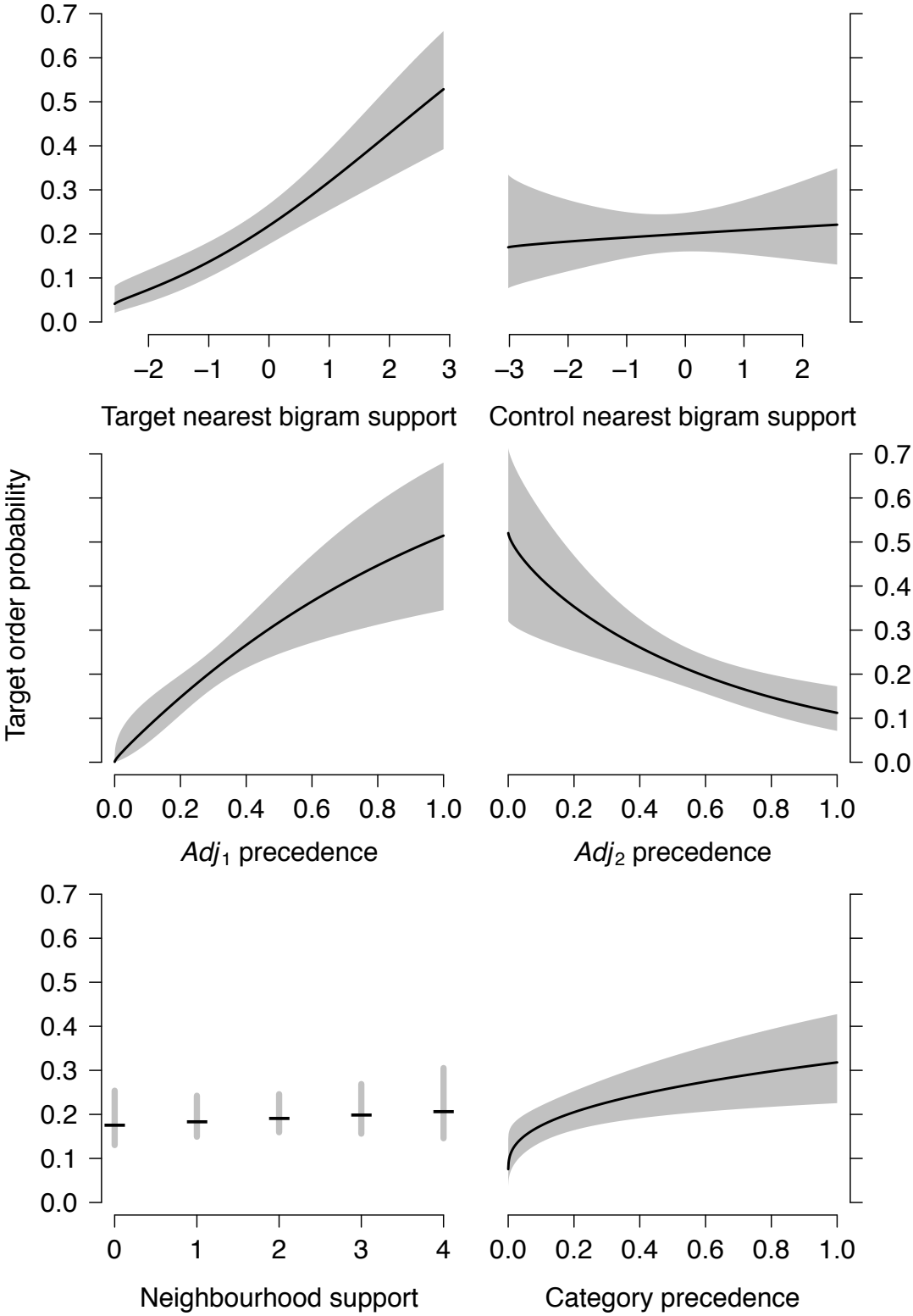

Figure 4. Partial effects of the Model 2 variables (see Table 4) on selected adjective order, at the median target order position of 0.5. The grey bands are approximate 95% confidence intervals.

The exemplar-based variables responsible for the reduction of the category precedence effect are adj_1 precedence and adj_2 precedence, as evidenced by the change in their coefficients from step one to step two of the hierarchical regression. This is not surprising, because those variables are explicitly meant to solve the same problem as category precedence, albeit using a different approach. All three variables capture the linear precedence relationship between two adjectives on the basis of the linear precedence relationship between abstractions over those adjectives. For category precedence, those abstractions are positional categories; for adj_1 and adj_2 precedence, they are the adjectives' neighbourhoods. This overlap between the category precedence and exemplar-based positional precedence is also evident from the fact that, of the four newly introduced exemplar-based variables, adj_1 and adj_2 precedence were the most strongly correlated with category precedence, as shown in table 5.

The target order's nearest bigram support seemed to cover a distinct part of the variation in selected adjective order. As opposed to adj_1, adj_2, and category precedence, this variable does not directly predict adjective order preferences. One expects the dispreferred order of an adjective pair to have a low nearest bigram support. However, a low nearest bigram support for a particular adjective order does not necessarily mean that the inverse order will have a high nearest bigram support, because the adjectives could simply be incompatible (e.g. *the wide narrow hallway* vs. *the narrow wide hallway*).

General discussion

The results of the 2AFC task and their reanalysis confirm the hypothesis that online abstraction over exemplars accounts for variation in adjective order preferences that category-based explanations fail to capture. The exemplar-based variables, which