DE GRUYTER
MOUTON

Dagmar Divjak,
Stefan Th. Gries (Eds.)

FREQUENCY EFFECTS IN LANGUAGE REPRESENTATION

TRENDS IN LINGUISTICS

Frequency Effects in Language Representation

Trends in Linguistics
Studies and Monographs 244.2

De Gruyter Mouton

Frequency Effects in Language Representation

edited by

Dagmar Divjak
Stefan Th. Gries

De Gruyter Mouton

ISBN 978-3-11-027378-6
e-ISBN 978-3-11-027407-3
ISSN 1861-4302

Library of Congress Cataloging-in-Publication Data

A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

Typesetting: RoyalStandard, Hong Kong
Printing: Hubert & Co. GmbH & Co. KG, Göttingen
♾ Printed on acid-free paper

Printed in Germany

www.degruyter.com

Preface

The work presented in this volume and its companion *Frequency Effects in Language Learning and Processing* (Gries & Divjak 2012) was originally selected for presentation at one of the two theme sessions that Stefan Gries and I organized for Summer 2009: a smaller theme session entitled "Converging and diverging evidence: corpora and other (cognitive) phenomena?" at the Corpus Linguistics 2009 conference in Liverpool (UK) as well as a larger one called "Frequency effects in language" planned for the International Cognitive Linguistics Conference at the University of California, Berkeley (USA).

This second volume draws theoretical conclusions from (mis)matches between different types of empirical data. Despite the importance attributed to frequency in contemporary linguistics, the link between frequencies of occurrence in texts and status or structure in cognition as reflected in experiments has not been studied in great detail, and hence remains poorly understood. Chapters in this volume explore the relationship between certain aspects of language and their representation in cognition as mediated by frequency counts in both text and experiment. They aim to contribute answers to questions such as: Which corpus-derived statistics correlate best with experimental results? Do certain types of corpus data fit certain types of experimental data better than others? Or, do corpus data have to be understood and analyzed in a radically different way to obtain the wealth of cognitive information they (might) contain?

To ensure quality volumes, every chapter was reviewed by three experts before the manuscripts were submitted to and reviewed by the series editors and their reviewers. We wish to thank our contributors for bearing with us through three review rounds, the reviewers for their efforts to ensure the quality of the contributions as well as the series editors and de Gruyter Mouton's Birgit Sievert, Julie Miess and Wolfgang Konwitschny for expediting the review and publication process.

Dagmar Divjak

Table of contents

Introduction

Dagmar Divjak

The frequency wars

How is it that we learn to communicate? The fundamental human capacity for acquiring and mastering language has long been of interest to linguists, psychologists, and neurologists alike. We take it for granted that any infant, in only a few years' time, will master at least the basics of a highly complex symbolic system. But how they accomplish the remarkable feat of learning a language remains a mystery. Not only does building artificial systems with the same capability for language remain out of reach despite decades of phenomenal advances in computing, it is far from obvious how it should be done.

This lack of knowledge about the way in which we acquire language has caused a rift between linguists of naturist and those of nurturist persuasion. Naturists believe that parts of the brain have evolved over time for the purpose of producing and understanding language. Most of the acquisition of natural languages by children depends on unacquired (or acquired but unlearned) linguistic knowledge or language-specific cognitive mechanisms while some features of natural languages are acquired by triggering. What is acquired in that case is a particular setting of an in itself unacquired parameter that specifies a fixed set of mutually exclusive linguistic properties, of which any given natural language can have only one. Grammar must be innate since the input children receive is considered ungrammatical and error-ridden (full of hesitations, self-corrections and ill-formed utterances), as well as highly complex, which constitutes a very poor database from which to deduct linguistic structure (cf. "poverty of the stimulus" debate). The research program of linguistic naturism or nativism therefore aims to show that very little knowledge of syntactic structure is acquired or learned from sensory stimuli. Instead, they hypothesize, infants are born with innate knowledge of how languages work and how they do not work. These innate biases allow children to figure out quickly, despite the poor quality of the limited amount of input they receive, what is and what is not possible in the grammar of their native language, and allow them to master that grammar by the age of three and a half.

Nurturists, on the other hand, claim that very little of the acquisition of natural languages by human beings depends on unacquired (or acquired but unlearned) linguistic knowledge or language specific cognitive mechanisms. Infants are born possessing a few basic cognitive abilities that develop through interactions with the environment but are independent of any inheritable code found in the genes. Acquiring a language is thus a process in which learning mechanisms, such as pattern discovery, that are a part of a general cognitive learning apparatus play a key role; this obviates the need for an innate and specialized language acquisition device. Over the last two decades, the nurturist approach to language has made impressive advances: following the discovery that children's path to language is gradual and piecemeal, still unfinished at the age of three and a half, and that adult speakers' language systems are sensitive to the frequencies of occurrence in language use, the conviction has grown that language is a dynamic system emerging from use. At least part of the solution to the problem of language acquisition may thus be found in input to which the learner is exposed.

Experiencing language as a messy "bag of words"

Human beings are quite adept at determining underlying frequency distributions and central tendencies. Research suggests that learners, even infants, can use statistical properties of linguistic input to discover structure, including sounds patterns, words and the beginnings of grammar (Saffran 2003, but see Yang 2004 who criticizes this work for its "unrealistic" setting). Exploiting this capability to extract regularities from sensory input turns language into a fine-tooled, acquirable system of communication. Therefore it seems reasonable to assume that probabilistic information and the ability to detect and abstract distributional characteristics of natural language that reflect underlying linguistic structure play a key role in linguistic development. Given this, one might hypothesize that what we learn is a probabilistic grammar grounded in our language experience rather than using environmental triggers to set parameters specifying a fixed set of mutually exclusive linguistic properties.

In such an experience-based grammar, the cognitive organization of someone's experience with language, linguistic categories and linguistic structures are associated with activation or probability values that are determined by their relative frequencies in language use (cf. Elman et al. 1996 among other classics). In attempts to understand the nature of the

mechanisms underlying language acquisition and representation, input-uptake correlations have the potential to provide deep insight: if particular aspects of input are predictive of language development and processing such findings shed light on the nature of language learning strategies and mechanisms, as well as the relationship of linguistic knowledge and experience.

Dealing with the effects of frequencies in language use on cognition is a discipline at the intersection of cognitive corpus linguistics and psycholinguistics. Yet many of the existing psycholinguistic studies on the theoretical import made by frequency have been carried out by research teams that do not include (corpus) linguists who have gotten their hands dirty in the data. Hence, many studies identifying frequency effects are based on analyses of a handful of phenomena, typically count readily identifiable forms considered independently from their wider linguistic context (which significantly diminishes the richness of the input from which human beings extract distributional patterns) and concern mostly one language, English. On the other hand, many cognitive corpus linguistic studies have taken their painstakingly annotated textual datasets to be a map of speakers' minds, forgetting that what is learned or acquired by probabilistic means is not strictly proportional to the stimulus (and that frequency is not the be all and end all in language, see Baayen 2010; Ellis 2012). Probabilistic learning theory holds that language learning is based on complex, higher-order properties of probabilistic patterns in sensory experience, not a mere tabulation of frequency of patterns (Elman 2003). Driven to its extreme, this split approach would reduce our billion-neuron brains that enable us to adapt quickly to an immense array of stimuli to nothing more than sophisticated abacuses used to keep tallies of all the words found in the messy bag of words that language is. We need to put the linguistics back into the frequencies and relate frequency distributions properly to the mind.

Over the last decade, the bulk of the research on frequency has focused on identifying frequency effects in an ever growing number of areas; frequency effects have now been documented in processing elements at about every level of traditional linguistic analysis, i.e., phonology, morphology, morphosyntax, syntax, lexicon (including formulaic language). It would seem we have ample proof of their existence. We also know quite a bit about the facilitatory role played by frequency distributions in breaking into the language system and building up the basic structures (see Ambridge & Lieven 2011). Time seems to have come to turn attention to less well-studied domains, for example, what kind of corpus-derived frequencies predict what kind of behavioural data best? Where are the

convergences and discrepancies between corpus and experimental data and what can those (dis)similarities teach us about the language system in the mind? Or to shift gears to theoretical questions that go beyond winning battles in the frequency wars: while frequency explains some of the variance, it does so in interaction with other determinants; how do these factors interact and how much of the cognitive structure that constitutes language in a mature speaker can be reduced to frequency (effects) alone (cf. Baayen 2010, Divjak in progress)? Such theoretical considerations give frequency the place it deserves in an encompassing model of usage and its effect on acquisition, systematization and representation.

Frequency and its effects coming of age

This second volume focuses therefore on the link between frequency and linguistic representation. It explores the relationship between well-studied aspects of language (constructional alternations, lexical contrasts and extensions as well as multi-word expressions) in a variety of languages (Dutch, English, Spanish and Russian) and their representation in cognition as mediated by frequency counts in both texts and experiments. The volume contains some contributions (Sokolova et al.; Schönefeld; Colleman & Bernolet; Littlemore & MacArthur) that rely on corpora to explore the richness of challenging constructional variation, capturing the linguistic frequency distributions in all their complexity. Other contributions (Teddiman; Snider & Arnon; Caldwell-Harris, Berant & Edelman) use advanced experimental techniques to investigate unanalzyed linguistic sequences from different frequency ranges while Van de Weijer and his collaborators combine both. State-of-the-art data collection (ranging from interviews to eye-tracking) and analytical techniques (from chi-squared tests to random effects regression) instill confidence in the theoretical conclusions drawn.

Constructional Alternatives and Alternations: corpus-based frequencies as diagnostic tool

The first three papers take a detailed look at constructional alternatives or alternations, with special attention to, on the one hand, the way in which corpus-derived frequencies can be brought to bear on the psychological reality of the patterns identified and, on the other hand, how measuring

verb-specific constructional preferences in a reliable way may resolve issues of divergence between corpus- and experimental frequencies.

Schoenefeld's paper is concerned with an empirical analysis of three types of English *go* complement$_{Adj}$ patterns, namely *go* plus *un*-V-*en* participles, *ing*-participles and original adjectives. She takes a usage-based construction grammar perspective using data from the British National Corpus. To learn about the special meanings and functions of the individual patterns and their interrelations, the data are analysed both quantitatively and qualitatively. More precisely, the data are submitted to collexeme and covarying collexeme analyses and – selectively – subjected to careful individual inspection. The results reveal that the three patterns are associated with different prototypical functions/meanings but, also, that more specific (that is, lexically filled) patterns can be recognized. In exploring how her data can be brought to bear on the psychological reality of the overall schematic, the partially schematic as well as the lexically filled patterns identified, Schönefeld looks at both type and token frequencies of the instantiations: these are known to affect entrenchment in different ways with a schema gaining strength (i.e., becoming 'established' or 'entrenched') in proportion to the type frequency of the instances which elaborate it, yet being weakened through the high token frequency of an instance.

Sokolova et al. present an empirical study that addresses critical aspects of two theoretical issues simultaneously, namely the nearly universally studied "Locative Alternation" and the specifically Russian phenomenon of allegedly "empty" aspectual prefixes. Their data, extracted from the Russian National Corpus, capture the behavior of the Russian verb *gruzit'* 'load', which participates in the Locative Alternation in both its unprefixed (*gruzit'*) and prefixed forms (*nagruzit'*, *zagruzit'* and *pogruzit'*). According to mainstream Russian linguistics, the prefixes *na-*, *za-* and *po-*, forming the prefixed counterparts of the verb *gruzit'* 'load', are considered semantically "empty", bearing only the aspectual feature "perfective". By fitting a logistic regression model to corpus data on the Locative Alternation, Sokolova et al. find that the four verbs behave differently in terms of the participating locative constructions they prefer, i.e., the Theme-Object vs the Goal-Object construction. While the unprefixed imperfective *gruzit'* favors the Theme-Object construction, *nagruzit'* strongly favors the Goal-Object construction, *pogruzit'* uses the Theme-Object construction virtually exclusively, and *zagruzit'* strikes a near-balance between the two constructions. Their findings thus support the hypothesis that the Locative Alternation requires the combination of three elements that each carry meaning, i.e., the prefix, the verb and the construction. The authors propose different

motivations for the observed frequency distribution, contributing to the ongoing discussion of what motivates the Locative Alternation.

Colleman and Bernolet focus on the crucial question of how to measure verb-specific constructional preferences in a reliable way. Their paper offers a detailed comparison of the results from a corpus-based investigation of the dative alternation in Dutch with the findings from a series of picture description experiments. This comparison reveals a striking contrast between both datasets in terms of the overall proportions of double object (DO) versus prepositional dative (PD) instances: whereas the DO construction is by far the most frequently realized option in natural language, the experimental data display a distinct bias toward the PD construction. The authors explore a number of factors contributing to this contrast, while showing that the alternation biases of the individual dative alternating verbs included in both investigations are quite consistent if they are measured appropriately. That is, if they are not simply measured in terms of the raw observed frequencies of DO and PD instances in both databases, but in collostructional terms, i.e., in a way which evaluates these observed frequencies against the frequencies expected on the basis of the overall distributions of the DO and PD constructions in the respective databases, while taking into account the fact that different senses of a polysemous verb may well exhibit different argument structure preferences.

Multi-word expressions: frequency and the privileged status of words in the mental lexicon

In the second part of this volume, research challenges the privileged position of words as prime units of mental representation and processing. Recently reports have been published that document frequency effects not only for single words, but also for sequences of two, three and even four words. Crucially, these frequency effects emerge not only for full phrases, idiomatic as well as non-idiomatic, but also for partial phrases (Tremblay 2009), and as the contributions in this section show, hold at all levels of the frequency spectrum, including levels of frequency so low the word "combinations" could be considered "random".

In their chapter, Snider and Arnon present experimental evidence challenging the distinction between 'stored' and 'computed' forms and supporting the single-system models by undermining the empirical criteria used to distinguish between the two models and demonstrating parallels in the processing of words and phrases, as well as of idiomatic and non-idiomatic

phrases. Data stem from the Switchboard, Fisher and British National corpora, norming experiments are performed over the web (on Amazon Mechanical Turk, www.mturk.com), reaction times studies use a phrasal decision and a sentence completion task and were analysed using mixed-model linear and logistic regression respectively. The results reveal that frequency affects the processing of 4-word compositional phrases at all frequency ranges: response latencies vary continuously across a range of low and high frequency four-word sequences. And idioms, while often thought to be holistically stored, show priming of their construction just like non-idioms.

Caldwell-Harris, Berant and Edelman set out to test two properties of the usage-based approach predicted by the ADIOS algorithm. In so doing they use frequencies from the Corpus of Contemporary American to analyze existing behavioral data from perceptual identification tasks, a paradigm which had not been previously used to study processing of multiword utterances. In a first study, they push Snider and Arnon's research a step farther by studying more of the frequency spectrum, demonstrating frequency effects for word pairs across the entire frequency continuum, from high frequency collocations over low frequency collocations and merely legal combinations to random word pairs, occurring on average only 1.3 times in the 410 million word COCA. Models which propose that statistics are maintained of exemplars stored only for sequences with some minimum frequency will find it difficult to account for the reported frequency effects. In a second study, Caldwell-Harris and Edelman find evidence for the prediction that language users who have more experience with specific linguistic stimuli exhibit more efficient processing of those stimuli. They support their claim with data drawn from an original source, i.e., Orthodox and secular Jews' processing of religious phrases, which have three frequency levels (daily, weekly and annual) and secular phrases, which have two frequency levels (common and rare). Compared to religious Jews, secular Jews have overall poorer identification of the religious phrases and show only weak frequency effects. This finding bears on the poorly understood frequency/entrenchment relationship that occupies a prominent position in corpus-based cognitive linguistics (cf. Schmid 2010).

With frequency typically being considered a marker of lexical storage, the number of lexical structures alone that needs storing is thus rapidly expanding beyond what a human brain can reasonably be expected to accommodate. This poses a serious threat to the cognitive plausibility of exemplar-based models to the extent that researchers are starting to

suspect that something else altogether may be going on (Baayen 2010), a conclusion that is echoed in the third and final part of this volume.

Lexical Contrasts and Extensions: frequency and the structure of the mental lexicon

The contributions grouped in this final part of the volume investigate the role frequency plays for the structure of the mental lexicon. More in particular, focus is on the relation between frequency of occurrence and knowledge of the many senses of polysemous items, between frequency of occurrence and awareness of lexical category membership for verbs and nouns affected by conversion, as well as between frequency of occurrence and goodness of antonymy.

According to the cognitive linguistic paradigm, word senses sit within radial categories with basic, often concrete, senses at the centre and abstract, figurative senses lying towards the periphery. In their chapter, Littlemore and MacArthur investigate the intuitions that both native and non-native speakers of English and Spanish have of the categories of senses associated with the words *thread*, *wing* and *hilar* ('thread') and compare these intuitions with findings from a corpus-based study; that study had identified significant, yet different, sense shifts between equivalent denominal verbs in the two languages. The results of the comparison show that, compared with the corpus data, the intuitive data for both native and non-native speakers are relatively impoverished and skewed. Compared with native speakers, even advanced learners have limited knowledge of the senses lying towards the periphery, and among the native speakers there is considerable variation, with younger speakers exhibiting different knowledge from older speakers. They conclude that radial category knowledge builds up over a lifetime, and even rich corpus data is unlikely to reveal the variable nature of category knowledge among individuals.

Teddiman focuses on the phenomenon of lexical conversion, a common and productive word-formation process in English that allows a single word to be associated with more than one lexical category without the use of overt morphological marking. She asks whether speakers of English are able to categorize categorically ambiguous words as nouns or verbs and whether the relative frequency with which words occur as nouns versus verbs influences the speakers' category choice. The results obtained in direct ratings tasks and a binary category decision task with educated native speakers of Canadian English suggest that this is indeed the case.

Frequency influences the storage of ambiguous items: speakers make use of their overall experience with a given lexical item in the absence of supporting context. Under this interpretation, words that undergo conversion are stored with probabilistic information specifying how likely a given word is to be encountered as a noun or a verb and this knowledge helps speakers distinguish between noun and verb reading when supporting context is missing.

Van De Weijer et al.'s study, finally, addresses the core of the frequency issue by investigating the role of frequency for strength of lexical affinity of antonym couplings in a methodologically exemplary way. In order to find out whether goodness of antonymy is a matter of strength of relatedness or lexical association and to what extent frequency of co-occurrence plays a role for the status of strongly canonical antonyms, they carry out a visual lexical decision experiment using materials based on frequency information obtained from the BNC, and analyse the data by fitting a multilevel model. Their findings confirm that antonymic targets are facilitated by their primes, but it casts doubt on the hypothesis that frequency of co-occurrence facilitates word recognition, either for antonyms or for unrelated adjectives. This means that there is a relatedness effect but no co-occurrence frequency effect, which indicates in turn that priming cannot be attributed to lexical association. The prime-target effect they obtain is a semantic effect, signalling that conceptual opposition is the *cause* of the lexical relation rather than the *effect* of the lexical relation, which lends support to a conceptual rather than a lexical approach to antonymy.

Selected references

Ambridge, Ben and Lieven, Elena
2011 *Child language acquisition: contrasting theoretical perspectives*. Cambridge: CUP.

Baayen, R. Harald
2010 Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon* 5(3): 436–461.

Divjak, Dagmar
(In progress) *Too rare to care? Frequency affects the acceptability of rare structures* (working title).

Ellis, Nick
2012 What can we count in language, and what counts in language acquisition, cognition, and use? In St. Th. Gries & D. Divjak (eds.), *Frequency Effects in Language Learning and Processing*. Berlin: Mouton De Gruyter. [Trends in Linguistics 244-1]

Elman, Jeffrey L., Elizabeth Bates, Mark Johnson, Annette Karmiloff-Smith, Domenico Parisi & Kim Plunkett
1996 *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press/Bradford books.

Elman, Jeffrey L.
2003 Generalization from sparse input. In *Proceedings of the 38th Annual Meeting of the Chicago Linguistic Society*.

Saffran, Jenny R.
2003 Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science* 12: 110–114.

Schmid, Hans-Jörg
2010 Does frequency in text instantiate entrenchment in the cognitive system? In Dylan Glynn and Kerstin Fisher (eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches*, 101–133. Berlin: Mouton de Gruyter.

Tremblay, Antoine
2009 *Processing Advantages of Lexical Bundles: Evidence from Self-paced Reading, Word and Sentence Recall, and Free Recall with Event-related Brain Potential Recordings*. Ph.D. Dissertation. University of Alberta, Edmonton, Canada. [Available for download at http://www.ualberta.ca/~antoinet/ThesisDraft_10_B.pdf]

Yang, Charles D.
2004 Universal Grammar, statistics or both? *Trends in Cognitive Sciences* 8(10): 451–456.

Things going unnoticed – A usage-based analysis of *go*-constructions

Doris Schönefeld[1]

Abstract

This paper is concerned with an empirical analysis of three types of *go* complement$_{Adj}$ patterns from a usage-based construction grammar perspective. The analysis is based on data from the British National Corpus (BNC), from which all instances of the verb *go* followed by an adjectival complement were extracted. The data fall into three formally different patterns, namely *go* plus *un*-V-*en* participles, *ing*-participles and original adjectives. To learn about the special meanings and functions of the individual patterns and their interrelations, the data were analysed both quantitatively and qualitatively. More precisely, the data were submitted to collexeme and covarying collexeme analyses and – selectively – subjected to careful individual inspection. The results reveal that the three patterns are associated with different prototypical functions/meanings and that there are also more specific (that is, lexically filled) patterns recognizable. The acknowledgement of the observed patterns as constructions is based on the patterns' observed type and token frequencies.

1. Introduction

This paper reports on an empirical analysis of three types of English *go* complement$_{Adj}$ patterns from a usage-based construction grammar perspective. Within this framework, it is assumed that native speakers' linguistic knowledge consists in their knowledge of constructions, or entrenched form-meaning pairs, which emerge from their linguistic experience (cf. Langacker 1987, 2000; Hopper 1987; Goldberg 1995, 2003, 2006; Barlow

1. The author would like to thank two anonymous reviewers and the editors of this book for extremely helpful comments and suggestions on earlier versions of this text, whose quality was greatly enhanced by their constructive criticism. Needless to say, the errors and inconsistencies remain my own. Author's correspondence address: schoenefeld@uni-leipzig.de

and Kemmer 2000; Bybee 2006; Bybee and Hopper 2001; Diessel 2007, for example). In order to recognize and inspect the respective *go*-patterns as constructions, their meaning potential and their interrelations, a quantitative corpus analysis is carried out. This is common practice in usage-based research, where potential constructions are assumed to be identifiable on the basis of what is frequent in speakers' performance as captured in corpus data. We hypothesize that the three *go* complement$_{Adj}$ patterns represent different constructions in that each is a symbolic (structural) configuration, i.e. a (complex) sign in which a particular form is paired with a particular function[2] (cf. Fillmore (1988: 36), Goldberg and Jackendoff (2004: 533)).

The patterns under analysis contain a form of the verb *go* followed by different types of adjectival complements, namely past participles with the prefix *un-* (so-called *un*-V-*en*), original adjectives and present participles, all of which can be used to express properties. The labels we use for the respective constructions are *go un*-V-*en* construction (1), *go* adjective construction (2) and *go* V-*ing* construction (3).

(1) (. . .) *war crimes in other parts of the world* ***go unpunished***.
(BNC, K5C)

(2) (. . .) *Whereas people* ***go bankrupt****, companies go into liquidation.*
(BNC, CD0)

(3) (. . .) *P. H. I've been on the street when helmets have* ***gone flying***.
(BNC, B24)

The analysis of the *go un*-V-*en* construction (example (1)) was informed by a study by Bourdin (2003), who characterized the construction semantically on the basis of impressionistic observations and a lexical search of the British National Corpus (BNC). In this paper, we revisit his semantic description of the construction and ask whether it holds up in light of the evidence drawn from an exhaustive search of the BNC, and how it relates to the other patterns that are similar in form. The formal similarity of the three constructions suggests that they are functionally/semantically related as well. More specifically, we assume that the constructions are located at the same level of schematicity in a network, (possibly) dominated by a more schematic '*go* complement$_{Adj}$' pattern. The more schematic pattern is expected to result from formal and semantic commonalities between

2. For a discussion of the notion of 'construction' within and outside a construction-grammar approach, see Schönefeld (2006).

the more specific patterns: all have the verb *go* followed by an adjectival complement and are all concerned with associating a property with a referent taking the subject position in the respective utterance. The (more specific) constructions are expected to be semantically distinct in that they (proto-)typically express different types of properties and particular associations between the subject referents and these properties. The quantitative analyses of the respective data are carried out to test these assumptions. On the one hand, the lists of attracted (covarying) collexemes contribute to the identification of each construction's function/meaning, while simultaneously making their interrelations visible. On the other, assuming that the entrenchment of cognitive units depends on usage, the type and token frequencies of the observed patterns are informative of their status as constructional units.

The study is organized as follows. Section 2 presents what is already known about the three *go*-constructions studied, section 3 informs about the methods used to analyse the corpus data, and section 4 reports on the results obtained. Section 5 is concerned with the discussion of the results and section 6 concludes the study by suggesting an inventory of constructions and specifying the relations assumed to exist between them.

2. *Go* complement$_{\mathbf{Adj}}$ patterns

2.1. The *go un*-V-*en* pattern

We use Bourdin's (2003) label *go un*-V-*en* construction for the type of construction illustrated in (1). Such expressions have also been addressed elsewhere, for example in Huddleston and Pullum's (2002) reference grammar. Describing adjectival passives with the negative prefix *un-* (often called "*un*-passives"), Huddleston and Pullum (2002: 1440) note that these special adjectives "can occur as complements to the verb *go* in the sense 'remain'". In earlier work, Hust (1977) investigates the unpassive construction in English from a generative perspective, characterizing it as containing the verb *be* and an *un*-participle, but also commenting on the use of the non-inchoative verb *go* in it (cf. Hust 1977: 38). In his discussion of the categorical status of the *un*-V-*en* forms, he argues that they are adjectives, since there is no verbal base (such as **undetect* for (1)). This view supports our assumption that the *go un*-V-*en* construction can be classed with the *go* adjective construction.

As for the historical development of the *go un*-V-*en* construction, we learn from the data collected in the OED[3] that the verb *go* is found:

23. In conjunction with adjs. having a negative sense, as *quit*, *unpunished*, *unrewarded*, etc. **where the original sense is that of leaving a court of justice or the like**, but passing now into that of continuing in a specified state (. . .). *to go free* (. . .).

(i) *a*1225 *St. Marher.* 18 Hit were þi gein þet tu þe
gest un ~ blescet. (. . .)
'it were your gain that to him
(you) go unblessed'
(OED Online, my emphasis)

The original sense (i) ascribed to this construction suggests that in the earliest uses, the *un*-V-*en* expressed the state in which the subject of *go*, the MOVER, was when leaving a scene. Such uses, (still) exhibiting a motional reading of *go*, have been described and analysed as one type of secondary predication, as depictives, the other type being resultatives[4] (cf. Aarts 1995; Rapoport 1990, 1999; Rothstein 2003, 2006, for example). The depictive subtype contains a "predicative attribute" which describes a characteristic of one of the participants associated with the main predicate. This characteristic is rendered as concomitant with the event encoded by the main predicate, which sets it apart from a resultative, where it is rendered as the result of the event (cf. Halliday 1967: 63; Himmelmann and Schultze-Berndt 2005: 4; Rothstein (2006: 210)). In the *go un*-V-*en* construction, the (predicative) attribute is ascribed to the subject participant, as in (4), adapted from the OED example (i):

(4) *They went unblessed.*
subject depictive – 'They (Subj) were unblessed when going/moving away'

This depictive use does, however, not play a part in Bourdin's analysis, who systematically excluded expressions in which *go* keeps its motional sense (as in X *go unaccompanied*) (cf. Bourdin 2003: 108). From the

3. We used the electronic version of the OED, last accessed on November 2nd 2010.
4. The terms depictive and resultative go back to Halliday (1967: 62). Both types "introduce a new event and define a relation between it and the event introduced by the main predicate." Rothstein (2003: 553–554)

analysis of his data[5] Bourdin (2003: 105–113) concludes that the nearest functional equivalent of *go* in the *go un*-V-*en* construction is almost always *remain*[6]. The state in which the subject remains is rendered by the *un*-V-*en* form and is construed as persisting over a period of time. The defining semantic feature of the construction as a whole is to express the notion of counternormativity/counterexpectation – crimes should, of course, be punished – and thus, quite naturally, using such an expression implies the speaker's negative judgement. The latter is assumed to be specifically conveyed by the verb *go*, an assumption that is in line with observations of the use of *go* in other constructions (for those see Quirk et al. (1985: 1174), Clark (1974: 316), and Radden (1996: 434), for example).

All things considered, the verb *go* in Bourdin's data can be said to have grammaticalized into a quasi-copula, comparable to that of *remain*. That is, the *go un*-V-*en* pattern represents two different constructions, the depictive and the (quasi-)copular construction, which we label attributive[7]. In the depictive construction (cf. OED example (i) and (4)), subject referents capable of motion trigger a motional reading of *go* and a depictive reading of the *un*-V-*en*. That is, we actually have two messages: the subject participant moves and is in a particular state. In the copular construction, (example (1)), the subject referent is incapable of motion and the only message of the utterance is the attribution of the state expressed by the *un*-V-*en* to the subject, hence the label attributive. The state is rendered as holding for a while rather than being merely ascribed to the subject. Contrasting this use with the *be un*-V-*en* construction, Bourdin (2003: 110) points at this feature of persistence, speaking of "differing aspectual profiles" of the two constructions. We think that this feature of continuation in the specified state is a natural effect of the use of the verb *go*, whose motional sense still reverberates in this use. The state ascribed to the

5. Bourdin's analysis is based on an empirical analysis of corpus data. Data were extracted from the BNC by means of lexical queries: all forms of the lemma GO were searched in collocation with 42 distinct verbal bases in the form of *un*-V-*en*, such as *unacknowledged* and *unused*. A complete list of the forms is, however, not made available.
6. This is in accordance with the OED's specification of 'continuing in a specified state' and with Huddleston and Pullum's observation (2002: 1440).
7. The term attributive is used by Hampe and Schönefeld (2003: 248–249, 2006: 135), though for constructions in which the property is predicated of an object, but it can analogously be transferred to the respective subject-related uses as exemplified by the *go un*-V-*en* instantiations and example (11) of the *go* adjective pattern.

subject is construed as moving through time, i.e. we have fictive motion, with the trajector being the state of the subject. The state's movement along the path is what triggers the notion of persistence. Hence, the resultant meaning is motivated by the image schema of motion[8] associated with the verb *go*[9]:

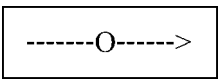


Figure 1. The image schema of motion (O = trajector ---> = landmark)

The two constructions also have some aspects in common: (i) the focus of the utterance is on the state expressed by the *un*-V-*en*, since depictive predicates also provide focal information (cf. Himmelmann and Schultze-Berndt 2005: 18), (ii) the state at issue is expressed by "stative-adjectival" past participles, which have generally been found to profile the final state of the process denoted by their verbal stems (Langacker 2008: 121–122), and (iii) the negative prefix *un-* contributes the notion of "absence of this (final) state", which is why the verbal process appears as not having taken place. However, despite these commonalities, the depictive and the attributive uses of *go un*-V-*en* have clearly distinct overall functions, so that we are concerned with two different, though closely related, constructions of the *go un*-V-*en* pattern. Our own analysis will take up this issue.

2.2. The *go* adjective pattern

In this section, we will turn to the *go* adjective construction. The complement in the *go* adjective construction takes the form of an original adjective, such as *bankrupt* and *crazy*. Also included are some deverbal adjectives, that is adjectival past participles which truly derive from verbs (vs the *un*-V-*en* already analysed), such as (*air-*) *borne*, (*self-*)*employed* and *tired*. This is done because they appear to denote properties which are less prominently profiled against the verbal processes they relate to, than

8. The idea that the image schema of motion is also involved in the attributive construal can furthermore be understood to motivate Bourdin's observation "that *go un*-V-*en* favours contexts involving temporal extension over those that focus on a moment in time." (Bourdin 2003: 113)
9. When the verb *be* is used, this notion of persistence of the state is absent, hence we have the construal of a state as it presents itself at the reference time, as in ***material*** *which* ***is unrecorded*** or *a* ***request was unanswered***, for example.

is the case for the *un*-V-*en* forms. Instantiations of the *go* adjective pattern comprise all functional patterns mentioned so far: depictives (example (5)), attributives (example (6)) and resultatives[10] (example (7)), as well as the lexically specific form *go wrong*, which is (almost) adverbial (example (8)).

(5) *If daughters Beatrice, . . . , and . . . Eugenie are there she does not* ***go topless****.* (BNC, CH6)

(6) *But first, Andrew Johnson explains why DOS is still* ***going strong*** (. . .) (BNC, FT8)

(7) *The landowners had not been paid, the company* ***went bankrupt*** (. . .) (BNC, HHG)

(8) *The problem of course arises, as I've said earlier, when things* ***go wrong***, (. . .) (BNC, F8N)

The potential adverbial tinge of the *go* adjective construction (noted for example (8)) seems to come from the notion of direction and the goal bias associated with the verb *go* (cf. Stefanowitsch and Rhode 2004: 250). Interestingly, the affinity with adverbials was found to be apparent in depictives, too (cf. Himmelmann and Schultze-Berndt (2005: 2, 4); van der Auwera and Malchukov (2005: 411), for example). The affinity is so high that German usage (9) does not even make a formal difference between what in English would be a depictive (10a) or an adverbial (10b).

(9)

Claire	*hat*	*wütend*	*das*	*Zimmer*
Proper[NOM.SG]	AUX[PRS PRF]	ADJ/ADV	ART	room[ACC.SG]

verlassen.
leave[PST PTCP]
'Claire has angry/ily the room left.'

10. Resultatives are usually object-related, as in *He wiped the table clean*, which roughly means 'the table (Obj) became clean as a result of V-ing'. It has, however, been observed in the literature that resultatives do relate to subjects of intransitive verbs, too. Levin and Rappaport Hovav (1995), among others, find that certain uses of intransitive verbs (such as *freeze*, *melt*, *break*, etc.), all being unaccusative verbs, require a resultative interpretation. In cases such as *The lake froze solid*, a property of the subject referent is predicated as a result of the event expressed by the verb: 'the lake (Subj) became solid as a result of freezing'. They attribute this observation to the fact that the surface subject of an unaccusative verb is the underlying direct object (cf. Levin and Rappaport Hovav 1995: 39)

(10a) *Claire left the room angry*.

(10b) *Claire left the room angrily*.

The problem of clearly distinguishing between these construction types has its repercussions in the classification of individual instantiations from the data, most obvious in the case of *X go wrong*. However, it does not speak against the existence of these patterns as different constructions, since there are instances in which the reading is unambiguous (as in (2), (5), and (6) above).

As for its historical development, the *go* adjective pattern is documented in the OED for the late 14th century, listed as sense 6 of the verb *go*. Semantically, it is specified for its attributive use.

6. With complementary adj. or equivalent phrase: To be habitually in a specified condition, esp. with regard to attire or circumstances affecting personal comfort. Now chiefly with reference to conditions implying neglect, privation, or disadvantage; (. . .)

(. . .) 1398 TREVISA *Barth. De P.R.* XVII. liii. (1495) 634 They **yede crownyd wyth iuy** that serued in the temple of Bachus. (. . .) *c*1511 *1st Eng. Bk. Amer.* (Arb.) Introd. 27 This people **goeth** all **naked**. 1535 COVERDALE *Gen.* xv. 2, I **go childles** (. . .).

Some of the examples cited are, however, also open to a depictive interpretation: 1398 *They went crowned with ivy* (. . .); 1511 (. . .) *this people goes naked*. The resultative use of the pattern is documented as sense 44:

44. To pass into a certain condition. Chiefly implying deterioration. a. With adj. complement: To become, get to be (in some condition). (Cf. COME 25a.) † to go less: to be abated or diminished. (. . .)

1583 T. STOCKER tr. Civ. Warres Lowe C. I. 117 The siege of Leyden continued, & their victuals **went** very **low**. (OED online, my emphasis)

The data given in the OED suggest that the attributive and resultative uses are later developments than the depictive.

2.3. The *go* V-*ing* pattern

The *go* V-*ing* pattern, illustrated in (11) and (12), is the third pattern included in the study, with the adjectival complement being represented by present participles.

(11) ***I** found myself laughing at how easy it all seemed, and axe in hand **went hurrying off** to join him.* (BNC, ECG)

(12) *Glancing over his shoulder with a huge wink, Oz banged the door after him and **went whooping** across the yard.* (BNC, AC4)

They are, however, different from the *un*-V-*en* participles in that they all relate directly to a verbal base and have a different semantic effect: They depict an activity in progress as a state. That is, they profile an (internal) portion of the event/activity with the successive states viewed as equivalent (cf. Langacker 2008, 120–121), the effect of which is to make an eventive predication stative (cf. Michaelis 2003: 194). Hence, the two participles in fact profile (aspects of) events as states. The denotation of a state is what these have in common with original adjectives (for a similar argument cf. Quirk et al. (1985: 168)). The individual forms differ in the internal make-up of the states expressed. Present participles prompt a stative construal of an internal portion of a process thus excluding its endpoint, past participles – a construal of its final state (cf. 2.1 above), and original adjectives – a construal of states that are depicted as unrelated to any process. This should also have an effect on the meanings and functions of the patterns in which they occur. Logically speaking, the *go* V-*ing* construction should be incompatible with a resultative reading, just like the *go un*-V-*en* construction. Yet the incompatibility in the former follows from the excluded endpoint of the event, whereas in the latter it is due to the explicitly negated final state. The *go* adjective construction is unconstrained in this respect and may take all functions discussed so far. As for the *go* V-*ing* construction illustrated in (11) and (12), we can specify its function as depictive. The *ing*-participles give information about the manner in which the subject moves (11), or point to what else the subject referent is doing when moving (12).

Instantiations containing *go shopping*, *go swimming* and the like suggest a different function, which is connected with the syncretism of the English present-participle ending and the verbal-noun suffix -ing. Historically, the V-*ing* form spread in Early Modern English as a form competing with *a*-V-*ing*, and both were used to render an action as being in progress (cf. Schlüter (2005: 209–210). Later on, *a*-V-*ing* lost ground against V-*ing*, which takes over also in the construction *go a*-V-*ing*. As a consequence of this development, Schlüter (cf. 2005: 228) notes a marked increase in the frequency of the *go* V-*ing* construction in the 20th century, when "the construction, which had been dominated by almost formulaic expressions like *go a begging/fishing/hunting/wooing* in earlier centuries, suddenly became

extremely productive and semantically flexible." As frequently found in grammaticalization research (cf. Hopper and Traugott 2001: 72), its high token frequency resulted in routinization, which also affected its meaning: the sequence lost its semantic transparency and *go* moved away from its other uses (cf. Bybee's (2006: 715) autonomy effect[11]). Thus, in this construction, *go* has been "reduced" or bleached to a quasi-auxiliary (cf. Schlüter 2005: 228). It signals the subject referent's movement away from the deictic centre with a focus on the GOAL component of the MOTION schema. Via the metaphor PURPOSES ARE DESTINATIONS, the expression thus refers to the subject referent's intended involvement in the process expressed by the V-*ing*[12]. That is, we have a purposive element superimposed, which is why we label this use of *go* V-*ing* as purposive. It now appears that the *go* V-*ing* construction can – just like the *go un*-V-*en* construction – be associated with two different functions, the depictive and the purposive, thus representing two constructions as well.

All in all, this amounts to quite crude assumptions: The *go un*-V-*en* pattern will be associated with the depictive and the attributive functions, the *go* V-*ing* patterns – with the purposive and depictive functions, and the *go* adjective pattern allows for any function except for the purposive, since the notion of an event is absent in adjectival complements. These selective observations and logical extrapolations need to be tested empirically, just as well as we need to determine how closely the constructions identified are associated with the three patterns under analysis, which would then allow for determining their interrelations. This is done by means of the quantitative analyses reported in the next section.

3. Corpus methods – a quantitative analysis of BNC data

The analyses reported here consider (in three consecutive steps) all occurrences in the BNC of all forms of the verb *go* immediately followed by *un*-V-*en* participles, original adjectives or present participles. It addresses questions of which functions/meanings the respective *go* patterns express in modern usage. In particular, it looks at the contributions of the adjectival complements, tests whether the patterns' functions are indeed as suggested in section 2, and investigates the patterns' interrelations.

11. Bybee (2006: 715) notes that frequently used sequences of words become more easily accessible as wholes, which reduces their analyzability "as they become autonomous from etymologically related forms (...)."
12. Such an element is also active in the grammaticalization path of *go* to *be going to* (cf. Bybee, 2006: 719–721, for example).

The required data were retrieved from the BNC using R. On the basis of scripts written by Stefan Th. Gries[13], we conducted a lemma search for the verb *go* preceding an element tagged as adjective, as past participle or as ambiguous between the two (the BNC tags being AJ0, VVN and AJ0-VVN[14]). The same was done for present participles. Since the BNC tags *ing*-forms following verbs as VVG (that is present participles), sometimes as VVG-NN1 (that is, a word-class ambiguous between a present participle and a singular noun, as after *enjoy*), the search for participles also extracts homonymous verbal nouns and cases ambiguous between an adjective and a present participle (that is elements tagged as VVG and AJ0-VVG). The extracted data were collected in separate output files for the respective patterns.

In order to semantically investigate the three patterns, a collexeme analysis was carried out. Such an analysis is assumed to provide "an objective approach to identifying the meaning of a grammatical construction and of determining the degree to which particular slots in a grammatical construction prefer, or are restricted to, a particular set or semantic class of lexical items." (Stefanowitsch and Gries 2003: 211). The slot considered is that of the adjectival complements (i.e. *un*-V-*en*, original adjectives and V-*ing*). To further elaborate on the meaning(s) of the patterns, the second variable/schematic slot in the pattern, its grammatical subject, is also included in the analysis, and the data were submitted to a covarying collexeme analysis.

These two methods belong to the *collostructional approach* developed by Gries and Stefanowitsch (cf. Stefanowitsch and Gries 2003 and 2005; Gries and Stefanowitsch 2004) for the investigation of the different kinds of associations between words and constructions on the basis of corpus frequencies. The former method relies on the (raw) frequency of a particular word in a particular construction, and on the (raw) overall frequencies of the word and the construction. From these frequencies, a measure of attraction/repulsion between a word and a construction is calculated, which is labelled collostruction strength. This measure has been shown to outperform raw frequency data (cf. Gries, Hampe and Schönefeld 2005: 635 and 2010: 71). In both production and comprehension tasks, it turned out to predict speakers' usage more exactly. For its calculation the Fisher-

13. I am much indebted to Stefan Gries for writing the scripts. I would like to emphasize here that I retain responsibility for further processing and interpreting the extracted data.
14. For more information on the BNC tags see http://www.natcorp.ox.ac.uk/docs/URG/posguide.html#m2.

Yates Exact test is used and the form in which the value is given is the negative logarithm to the base of ten of the *p*-values[15]. The latter method investigates the interactions between words occurring in two different slots of the same construction, extrapolating "whether and how different slots in a construction are related semantically." (Stefanowitsch & Gries 2005: 11). The significance of the relevant co-occurrence frequencies, called the covarying collexeme strength, is again calculated by means of the Fisher-Yates Exact test, drawing on frequency data for the two words in the two constructional slots to be investigated and all other words encountered in the same slots of the construction[16]. The type of analysis carried out here is item-based, i.e., covariation is analysed in a subcorpus of the total of instances of the respective pattern, the overall corpus frequencies are thus neglected (cf. Stefanowitsch and Gries 2005: 23). The slots considered in the *go un*-V-*en* construction are those of the subjects and the *un*-V-*en*. This makes sense because i) subjects were found to be involved in triggering the different readings of the verb *go*, and ii) the subject slots in the *go un*-V-*en* (or *un*-passive) construction denote AFFECTED referents (rather than AGENTS) and should, hence, be assumed to co-vary with the *un*-V-*en* participles to the same extent as transitive verbs and their objects (i.e. processes and AFFECTED participants) do[17]. This is in contrast to prototypical active sentences, where the subject's role is AGENT. As AGENTS – unlike AFFECTED participants – exist independently of the event expressed by the verb (cf. Goldberg 1995: 116), these do not necessarily enter into systematic covariation with the verbal predicate.

The investigation of the other two patterns proceeded analogously.

4. Results of the corpus studies

Section 4 presents the results obtained from the analyses of the BNC data for the individual patterns.

15. For a detailed elaboration of the method see also Gries, Hampe and Schönefeld 2005.
16. Since these methods are by now well-known in (quantitative) corpus linguistics, we abstain from describing them in more detail and refer the reader to the literature cited above, if further information is required.
17. The grammatical relation between the subjects and the process is also hinted at by Hust (1978: 79): "Unpassive participles select as their subjects the same class of nouns that the related active verbal stems take as objects."

4.1. The *go un*-V-*en* pattern

The R-based search of the BNC resulted in a total of 541 tokens of the *go un*-V-*en* construction, representing 107 types.

4.1.1. Collexemes

The collexeme analysis[18] revealed which *un*-V-*en* are most strongly attracted to the construction, which are less so, or even repelled. The 40 most strongly attracted *un*-V-*en* are given in Table 1, listing the attracted *un*-V-*en* participles in the order of collostruction strength, as well as showing their frequencies in the corpus (word freq) and the number of occurrence in the pattern (obs. freq).

Table 1. 40 most strongly attracted collexemes of the *go un*-V-*en* construction in the BNC

rank	word	word freq	obs. freq	coll. strength[19]	rank	word	word freq	obs. freq	coll. strength
1	unnoticed	396	134	Inf	21	unused	491	5	6.77
2	unchallenged	214	51	135.71	22	unfilled	61	3	6.30
3	unpunished	44	31	101.80	23	unobserved	89	3	5.80
4	unheeded	65	32	97.60	24	undiscovered	101	3	5.64
5	unrecognized	176	27	66.28	25	unregarded	11	2	5.49
6	unreported	130	23	58.10	26	untended	19	2	5.00
7	unremarked	41	16	46.96	27	untaxed	25	2	4.76
8	undetected	129	18	43.60	28	unasked	32	2	4.54
9	unanswered	233	19	41.32	29	untreated	245	3	4.49
10	unrecorded	84	13	32.30	30	unexamined	34	2	4.49
11	unrewarded	26	9	26.09	31	uncollected	37	2	4.41
12	unheard	202	11	22.21	32	unread	50	2	4.15
13	unchecked	150	9	18.68	33	untested	79	2	3.75
14	unmentioned	17	5	14.30	34	untouched	474	3	3.64
15	unstated	72	5	10.95	35	unactioned	1	1	3.62
16	undiagnosed	25	4	10.37	36	unpestered	1	1	3.62
17	unquestioned	95	5	10.34	37	unrevived	1	1	3.62
18	unseen	486	7	10.29	38	unsung	100	2	3.55
19	unsatisfied	92	4	8.03	39	unmet	120	2	3.39
20	uncontested	36	3	7.00	40	unreplaced	2	1	3.32

18. The collostructional analyses drawn on in this paper were computed with an interactive programme, Coll.analysis 3.2, an R script written by, and freely available from, Stefan Th. Gries.
19. Coll. strength: -log (Fisher exact, 10); ($>3 \Rightarrow p < 0.001$)

It turned out that there are no *un*-V-*en* collexemes repelled by the pattern. This does not really come as a surprise, because such participles have generally been discussed as occurring in the so-called unpassive construction following the verbs *be* and *go* (cf. Hust 1077: 38; Huddleston and Pullum 2002: 1440). The significantly attracted *un*-V-*en* collexemes are revealing of the pattern's semantics. Firstly, the base forms of the collexemes, that is, the unprefixed verbs (*notice*, *challenge*, *punish* etc.), can be classified into a number of semantic groups (taken from Levin's (1993) categories of English verbs). These are informative about the events whose absence is communicated by the *un*-V-*en* construction. Within the top 40 collexemes the following groups show up (given in the order of the collostruction strength of the item most closely associated with the construction):

- verbs of perception: *notice* (rank 1), *recognize* (5), *detect* (8), *hear* (12), *see* (18), observe (23) etc.
- verbs of psychological state/activity: *challenge* (2), *punish* (3), *heed* (4), *reward* (11), *satisfy* (19) etc.
- verbs of communication: *report* (6), *remark* (7), *answer* (9), *mention* (14), *state* (15) etc.
- image-creation verbs: *record* (10)
- search verbs: *check* (13), *examine* (30), *test* (33)
- verbs describing properties of events (characterize verbs): *diagnose* (16), *use* (21), *treat* (29)
- others: *tax* (27) – measure verb; *touch* (34) – verb of contact; etc.

Being associated with AFFECTED participants, the verbs generally denote bounded (telic) events, i.e. accomplishments and achievements (cf. Vendler's (1957) aspectual classes of verbs). More specifically, the verb bases identified allow for a more specific semantic description of the construction. The findings suggest that the construction is not freely available for signalling the absence of any bounded (telic) event, but is preferred for bemoaning the absence of events of perception, psychological activity and communication[20].

20. An anonymous reviewer suggested that the construction typically comes with an existential presupposition, which constrains the set of verbs that can occur in the construction to those taking 'affected' objects. It is true that the majority of object-participants is affected indeed – rather than effected, created by the verbal process that is. However, as examples of *books and papers that go/went unwritten* (revealed by a web search) suggest, this is not definitional for the verbs used in this construction.

A further aspect to consider for the semantic characterization of the pattern is the collexeme *unpunished*, ranking 3rd. Considering that this collexeme is quoted as the origin of the depictive construction, one may assume that depictives play a central part in the pattern. However, it turns out that for the functional specification of the pattern the attracted adjectival collexemes are not really informative. Some *un*-participles are easily compatible with a depictive reading (where we have movement and the ascription of a state to the subject): *unnoticed*, *unchallenged*, *unpunished*, *unrecognized* etc., they, however, do not exclude an attributive reading. Other collexemes do not suggest a depictive reading in the first place, such as *unanswered, unremarked, unstated, uncollected* etc. That is, though there are tendencies to recognize, the pattern's function cannot be determined clearly from the instances of *un*-V-*en* alone.

4.1.2. Covarying collexemes

The covarying collexeme analysis was expected to shed further light on the semantic-functional specification of the *go un*-V-*en* construction. It was used to detect whether there is any systematic covariation to be found between the *un*-V-*en* participles and the pattern's subject referents as specified in section 3.

The analysis of the 541 tokens results in 482 types of co-variation against 107 types for the collexemes, that is, the constraints for a particular *un*-V-*en* to occur in the *go* pattern seem to be stronger than for its co-occurrence with particular subjects. Table 2 gives the top 20 attracted covarying-collexeme pairs in the *go un*-V-*en* construction, as well as the words' corpus frequencies (freq.w1, freq.w2) and their observed (obs.w1_2 in_c) and expected (exp.w1_2 in_c) frequencies of co-occurrence. Note that the subjects have been lemmatized in the analysis. The lower end of the list, containing the least attracted and repelled covarying-collexeme pairs, is given in the appendix (Table A).

As Table 2 shows, only the first 11 combinations reach high significance ($p < 0.001$); but more than half co-vary significantly ($p < 0.05$), i.e., they co-occur more often than chance would predict. Three of the pairs are found to be in a relation of repulsion, that means if word 1 is used word 2 must be considered as quite unlikely to occur in the construction (cf. Table A). We note in passing that *unnoticed*, the top-ranking collexeme of the construction, appears in only few significant combinations with subjects at lower ranks, such as: *entry* (rank 152) and *presence* (153) (cov.coll.strength = 1.83). This may be due to the diversity of things

Table 2. 20 most strongly attracted covarying-collexeme pairs in the *go un*-V-*en* construction in the BNC

rank	word 1	word 2	freq. w1	freq. w2	obs. w1_2 in_c	exp. w1_2 in_c	relation	covarying collexeme strength[21]
1	untreated	tumour	3	3	3	0.02	attraction	7.42
2	unheeded	warning	32	5	5	0.3	attraction	6.27
3	unanswered	plea	19	6	4	0.21	attraction	4.80
4	unmet	need	2	4	2	0.01	attraction	4.39
5	unheeded	call	32	9	5	0.53	attraction	4.25
6	unreported	case	23	3	3	0.13	attraction	4.17
7	unasked	question	2	5	2	0.02	attraction	4.16
8	unanswered	letter	19	4	3	0.14	attraction	3.84
9	unchecked	rising	9	2	2	0.03	attraction	3.61
10	unheard	voice	11	2	2	0.04	attraction	3.42
11	unchallenged	statement	51	3	3	0.28	attraction	3.10
12	undetected	error	18	2	2	0.07	attraction	2.98
13	unaccompanied	staff	1	1	1	0	attraction	2.73
14	unactioned	complaint	1	1	1	0	attraction	2.73
15	unbeaten	we	1	1	1	0	attraction	2.73
16	unchastened	impudence	1	1	1	0	attraction	2.73
17	unclaimed	pence	1	1	1	0	attraction	2.73
18	uncompensated	loser	1	1	1	0	attraction	2.73
19	unconfirmed	rumour	1	1	1	0	attraction	2.73
20	unconsidered	objection	1	1	1	0	attraction	2.73

that can be noticed. That is, we do not seem to associate very specific scenarios with the act of noticing, and the large number of subjects whose association with *go unnoticed* does not reach significance is indicative thereof.

More generally, it turns out that the significant combinations are related to typical (transitive) events, which gave rise to the emergence of the respective collocations: *treat tumours*, *heed warnings*, *report cases*, *answer pleas*, *meet needs*, *answer letters*, *detect errors* etc., some of which are also frequent in passive construals (*treated tumours*, *reported cases*, etc.).[22] Secondly, the covarying collexemes show a large number of inanimate

21. Covarying collexeme strength: -log (Fisher exact, 10) ($>1.3 \Rightarrow p < 0.05$)
22. This finding confirms Stefanowitsch and Gries' observation (2005: 34) that the words occurring in different slots of a construction may exhibit semantic coherence on the basis of world knowledge as organized in frames.

subjects, which is indicative of the predominance of an attributive function of the pattern. The correlation is, however, not that simple, as it is not only animate subject referents which may trigger a depictive reading. Some inanimate subjects also permit a depictive construal. These are (i) products of communication, whose co-occurrence with particular *un*-V-*en* participles is significant: *warning* (rank 2), *plea* (3), *call* (5), *question* (7), *letter* (8), *statement* (11), *complaint* (14), *rumour* (19), and (ii) percepts, which rank lower, but also covary more often than chance would predict: *voice* (rank 10), *finds* (142) (cov. coll. strength = 1.88). Depending on the transience or permanence of the subject referent, they trigger either depictive or attributive readings. Transient referents of the first type, such as *warnings* (*going unheeded*), are involved in the construal of a scene of metaphorical motion. The warnings, the sounds of the words expressing them, fade away, and, since they were not heard, did so in the state of being unheeded[23]. More permanent referents, such as *letters* (*going unanswered*), signal a persistent state. The referents to which a certain state is ascribed do not disappear, that is they give *go* a clear attributive reading. The same holds true for the second subject type: *Voices* (*going unheard*), a more transient referent, triggers the metaphorical motion reading, i.e. a depictive, *finds* (*going unrecognized*), a more permanent referent, triggers an attributive reading. Apart from that, prosodic features may be drawn on for the identification of a depictive or attributive reading: Depictives could be given two intonational contours, whereas attributives never would. Information on that is, however, unavailable in the data. A more easily applied test relates to Bourdin's characterization of *go un*-V-*en* as a functional equivalent of *remain*, that is, all the attributive uses can be substituted by *remain*, whereas depictives cannot.

To sum up, the data show that in modern usage, the *go un*-V-*en* pattern realizes two distinct functions, so that from our understanding of a construction as a form-function pairing, it actually represents two constructions, the attributive and the depictive *go un*-V-*en*. Taking into account the historical relationship between the two, we would assume them to be related by constructional polysemy (for an argument about this topic cf. Goldberg 1995: 31–39). More precisely, the attributive construction can be understood as a grammaticalized form (via metaphor) of the depictive construction, with the mappings as suggested in section 2.1 above. From

23. This can be motivated by the metaphor EXISTENCE IS PRESENCE or, in one of its logical deductions, DISAPPEARANCE IS MOTION AWAY, and the metonymic reading of the subject (FORM STANDS FOR CONTENT)

the perspective of modern English usage, the attributive construction has surpassed the depictive and represents the more typical use.

4.2. The *go* adjective pattern

The BNC query returned 4,545 instances of the pattern '*go* adjective', representing 402 types (including 78 tokens/34 types of past participles). Again, these data have been submitted to frequency counts and a collexeme analysis in order to find the most strongly attracted adjectival collocates of *go* in this pattern and the meaning(s) associated with the construction.

Table 3. Most strongly attracted collexemes of the *go* adjective construction in the BNC

rank	word	word freq	obs. freq	coll. strength	rank	word	word freq	obs. freq	coll. strength
1	bust	168	161	Inf	21	bonkers	43	13	30.52
2	mad	2965	385	Inf	22	deaf	2629	34	30.50
3	wrong	9375	1333	Inf	23	pale	3219	36	30.09
4	bankrupt	503	137	307.27	24	strong	15441	66	29.54
5	berserk	121	87	245.09	25	barefoot	219	17	28.81
6	crazy	1698	118	189.24	26	cold	6814	44	26.88
7	left[24]	507	74	143.94	27	extinct	431	18	25.54
8	hungry	1786	71	96.69	28	dry	5271	38	24.98
9	quiet	5482	67	57.68	29	pink	2433	28	23.92
10	blank	1248	41	52.82	30	grey	4139	32	22.06
11	blind	2277	46	49.57	31	rigid	1407	22	21.78
12	sour	597	32	48.26	32	barmy	89	11	21.21
13	soft	5868	58	45.00	33	native	2333	25	20.72
14	numb	238	23	40.97	34	senile	170	12	20.05
15	bald	611	28	40.40	35	scarlet	571	15	18.40
16	white	18794	79	34.71	36	live	2532	23	17.58
17	dead	10873	62	34.52	37	brown	3880	26	16.61
18	mouldy	108	17	34.30	38	insane	363	12	16.05
19	haywire	24	12	31.49	39	overdrawn	76	8	15.02
20	free	19461	76	31.39	40	weak	3477	23	14.67

24. The form *left* in *go left* is erroneously tagged as adjective in the BNC. That is why it occurs in this data set, although it is clearly adverbial, namely indicating direction.

4.2.1. Collexeme analysis

On the basis of the most strongly attracted collexemes, we can identify two predominant functions (or meanings) of the pattern: (subject-) resultative and (subject-) attributive, with the states predicated of the subject referents construed accordingly. The former group is represented by such instantiations as *go bust* and *go mad* (ranks 1, 2), the latter – by *go hungry* and *go strong* (ranks 8, 24). In a few of the latter cases, the subject referent may additionally be depicted as moving away from a location, as in *go free*, *go barefoot* (ranks 20, 25), resulting in a depictive reading. A second noteworthy fact is the relatively high token frequency of the top collexemes in the pattern. This suggests that such expressions are known by the native speaker as collocations with their respective readings. Thus *go mad*, for example, is usually understood as resultative, although, technically speaking, a depictive reading is not inconceivable.

At the lower end of the list, the analysis unearthed 31 collexemes that are significantly repelled by the construction, such as *good*, *different*, *special* and *low* (cf. Table B (appendix)). Explanations saying that *go* in resultative readings signals deterioration, or departure from a normal state (cf. section 2.1) can account for some of them only. Most of the forms, it seems, are more at home in other types of constructions and their usage in the *go* adjective pattern must be considered accidental.

4.2.2. Results of the covarying collexeme analysis

Analogously to the previous construction, the pairs of covarying collexemes were checked for their contribution to a more specific functional description of the *go* adjective construction. The most strongly attracted pairs are listed in Table 4.

The top ranking covarying subject referents seem to indicate a preference for inanimate subject referents, which, however, disappears if all attracted subjects are included in the discussion. There are many pronominal subjects (*he*, *she*, *I*, *you*, *everybody* etc.) and proper names, so that the *go* adjective construction is open to messages about both animate and inanimate things. Just as noted for the *go un*-V-*en* construction, the type of subject referents is prominently involved in the sense the construction is going to render: animate referents are hospitable to literal-motion senses and metaphorical (change-of-state) readings. Inanimate referents are more at home in such change-of-state readings in which intension and agentivity are backgrounded. On the other hand, *go* may add some agentive notion to

Table 4. 20 most strongly attracted covarying-collexeme pairs in the *go* adjective construction in the BNC

rank	word 1	word 2	freq. w1	freq. w2	obs. w1_2 in_c	exp. w1_2 in_c	relation	cov. coll. strength
1	wrong	things	1333	332	313	97.37	attraction	150.96
2	wrong	what	1333	204	203	59.83	attraction	111.02
3	left[25]	you	74	106	57	1.73	attraction	84.77
4	wrong	something	1333	153	148	44.87	attraction	73.33
5	wrong	anything	1333	92	91	26.98	attraction	47.62
6	dry	mouth	38	27	23	0.23	attraction	47.26
7	dead	line	62	23	22	0.31	attraction	41.50
8	bust	company	161	59	32	2.09	attraction	31.46
9	blank	mind	41	19	15	0.17	attraction	28.35
10	mad	I	385	277	84	23.46	attraction	27.35
11	white	face	79	51	21	0.89	attraction	24.31
12	dead	phone	62	14	13	0.19	attraction	23.68
13	bust	firm	161	40	22	1.42	attraction	21.72
14	bankrupt	company	137	59	20	1.78	attraction	16.01
15	bust	business	161	15	12	0.53	attraction	14.97
16	mad	he	385	237	59	20.08	attraction	14.42
17	close	proper	11	293	11	0.71	attraction	13.17
18	wrong	nothing	1333	24	24	7.04	attraction	12.85
19	fine	everything	20	91	10	0.4	attraction	12.01
20	quiet	it	67	242	22	3.57	attraction	11.93

the inanimate subject referents, so that the things changing can be construed as playing a quasi-active part in the change. This is shown in example (13).

(13) *Anne heard her **voice go heavy** on the last word, as if it came from deep within her, influenced by some deep-felt experience.* (BNC, CCD)

In addition to these general observations, the covarying collexemes reveal some interesting more specific combinations: we find that the adjective *wrong* systematically associates with *thing|s*, *what*, *something*, *everything*, *nothing* in the subject slot, *dry* – with *mouth|throat*, *dead* – with *line|phone*, *blank* – with *mind|screen*, *white* – with *face|knuckles*, *bust* – with *company|firm|business*. The collexeme pairs identified corroborate the

25. See the previous footnote.

observations made for the *go un*-V-*en* construction: the associations between the two slots can be considered as reflecting a typical relation from our experience with the world. The frame-based association between the two words is even more straightforward: the states denoted by the adjectives are typically related to the subject referents, with which they equally typically co-occur in adjective-noun patterns. However, though we are concerned with typical state ascriptions, most of the significantly co-varying adjectival collexemes show that the state is construed as "departing from a normal state", an effect attributed to the verb *go* (cf. Clark 1974; Radden 1996).

In sum, the analyses of the pattern *go* adjective show that it is polysemous, too. "[T]he same form is paired with different but related senses" (Goldberg 1995: 33), representing the resultative, the attributive and – in a few cases – the depictive construction. The ranking of the adjectival collexemes suggests the resultative to be the prototypical construction.

4.3. The *go* V-*ing* pattern

The frequency, productivity and semantic flexibility attested for the *go* V-*ing* construction in the 20th century (cf. section 2.3) is also reflected in the data extracted from the BNC. We identified 395 (participle) types instantiated by 1767 tokens of the pattern, which we submitted to the same types of analyses as the other two patterns.

4.3.1. Results of the collexeme analysis

Again, the analysis reveals collexemes attracted to and repelled by the construction. The top 40 attracted collexemes are listed in Table 5, the latter are exemplified by *going*, *sitting*, *seeing* (cf. Table C in the appendix).

It is obvious that the top-ranking collexemes predominantly realize the purposive sense of going out in order to pursue an activity or a job (cf. also Schlüter 2005: 222). Due to the attenuated sense of *go*, the focus of the message is on the intended activity, the expression as a whole signaling that the subject referent is occupied with it:

(14) *There is a very definite distinction between those who* ***go climbing*** *and those who* ***go walking***. (BNC, K5D)

The verb bases of the *ing*-forms found mainly denote intransitive and atelic events. Quite a few name popular spare time activities, such as (*go*) *shopping*, *fishing*, *swimming*, *hunting* etc. The noticeable frequency of occurrence in this pattern suggests that the complete expressions are

Table 5. 40 most strongly attracted collexemes of the *go* V-*ing* construction in the BNC

rank	word	word freq	obs. freq	coll. strength	rank	word	word freq	obs. freq	coll. strength
1	shopping	587	293	Inf	21	racing	1081	17	12.98
2	fishing	130	80	196.73	22	sprawling	9	5	12.43
3	missing	1212	121	184.36	23	crashing	276	10	11.41
4	swimming	506	88	156.60	24	bowling	69	7	11.33
5	hunting	326	43	71.25	25	gallivanting	16	5	10.89
6	skiing	139	31	59.28	26	cleaning	756	12	9.42
7	marching	226	31	52.14	27	jogging	70	6	9.34
8	camping	68	22	46.38	28	storming	75	6	9.16
9	dancing	600	27	32.10	29	running	6616	31	9.06
10	rushing	700	27	30.30	30	visiting	1024	13	8.98
11	flying	1120	31	30.30	31	surfing	86	6	8.80
12	sightseeing	14	10	26.06	32	hill_walking	3	3	8.71
13	looking	24825	100	22.88	33	horse_riding	3	3	8.71
14	sailing	530	20	22.46	34	rafting	3	3	8.71
15	wandering	625	20	21.05	35	boating	15	4	8.49
16	begging	387	17	20.32	36	charging	318	8	8.02
17	riding	1530	24	18.00	37	rabbiting	21	4	7.85
18	shoplifting	12	7	17.44	38	walking	3498	20	7.44
19	chasing	681	17	16.23	39	rowing	37	4	6.81
20	haring	18	6	13.17	40	canoeing	9	3	6.79

candidates of lexically filled constructions. That is, they represent collocations reflecting people's habitual activities and other typical scenarios. Another group of collexemes triggers a depictive reading (*marching*, *rushing*, *running*), given the appropriate context. In both uses, the *ing*-form, focusing on an internal portion of the process, adds the notion of ongoing process.

In addition to that, there are quite a few collexemes on the list which deserve individual discussion: *missing*, *looking*, *begging*, *troubling*, *using*, for example. They represent a rather mixed bag of semantic verb classes (denoting bounded and unbounded events), so that they do not easily, or not at all, unify with one of the functions discussed so far. From Bourdin's analysis (2003: 114), which also looks at specific *go* V-*ing* patterns, we learn that these express "more grammatical functions" such as "disapproval" by the speaker[26] or a more aspectual meaning, namely that of inchoation, continuation and/or iteration.

26. The same specification is also given by Goldberg (2006: 53).

4.3.2. Results of the covarying collexeme analysis

In order to see whether such functions can also be identified for the data extracted in our analysis, this section looks at the pattern's (significantly) covarying collexemes, that is, the full expressions including their subjects, and their wider contexts.

Table 6. 20 most strongly attracted covarying-collexeme pairs in the *go* V-*ing* construction in the BNC

rank	word 1	word 2	freq. w1	freq. w2	obs. w1_2 in_c	exp. w1_2 in_c	relation	cov. coll. strength
1	marching	saints	30	23	22	0.39	attraction	42.14
2	begging	chance	18	4	4	0.04	attraction	8.11
3	swimming	ducks	88	6	6	0.3	attraction	7.87
4	shopping	I	293	224	67	37.38	attraction	7.09
5	missing	money	121	6	6	0.41	attraction	7.02
6	buzzing	bee	2	2	2	0	attraction	6.19
7	begging	jobs	18	3	3	0.03	attraction	6.04
8	missing	anything	121	4	4	0.28	attraction	4.67
9	shoplifting	boy	7	3	2	0.01	attraction	4.39
10	begging	wits	18	2	2	0.02	attraction	4.00
11	beating	unclear	2	28	2	0.03	attraction	3.61
12	marching	soul	30	2	2	0.03	attraction	3.55
13	getting	you	10	300	7	1.71	attraction	3.52
14	missing	artworks	121	3	3	0.21	attraction	3.49
15	missing	files	121	3	3	0.21	attraction	3.49
16	battering	monsters	1	1	1	0	attraction	3.24
17	beating	readers	1	1	1	0	attraction	3.24
18	bellowing	round	1	1	1	0	attraction	3.24
19	clutching	claws	1	1	1	0	attraction	3.24
20	dragging	objects	1	1	1	0	attraction	3.24

The top ranking collexeme pair, *saints* and *marching*, stands out with the highest collostruction strength value by far. However, this is due to the song *When the saints go marching in*, which accounts for all the cases of co-occurrence in the data. Hence, it can be excluded from consideration here, the more so since it does not contribute a new sense.

For the other collexeme pairs, we firstly note a correlation between the purposive reading and subject referents capable of (self-propelled) motion.

The same holds true for the depictive (manner) sense. Both readings presuppose their subject referents to have intentional minds, an aspect that becomes even more obvious in the first sense, where the subject referent moves purposefully. Secondly, we have a number of inanimate subject referents which normally do not trigger a motional reading of *go*, such as *chances|jobs|wits going begging, money|artworks|files going missing*, and *objects going dragging* (*along*). Most of these instances express what could be called "agentless motion"[27]. The inanimate subject referents are construed as PATIENTS (of motion), so that the message is that they "moved away" with the causer of the motion demoted. This could possibly be motivated by the fact that the speaker simply does not know the agent (as in example (15)), or wants to keep the agent hidden (as in example (16)).

(15) *They said there'd been a fire, some of the* ***files had gone missing*** *years back*. (BNC, HDC)

(16) (. . .) *it appeared that Abingdon could not turn possession into goals as* ***chance after chance went begging***. (BNC, KS7)

Hence, these constructions compare to ergatives. The instigator of the action is blanked out, we are just informed about what happened to the PATIENTS in an event, which surface as the subjects in the construction. The scene is construed as ongoing, an effect brought about by the PATH image schema of the verb *go* and the form of the participle (cf. section 2.3). In this respect, the expression in (15) is similar to *get lost*, which triggers a more dynamic reading than the normal passive *be lost*, but puts emphasis on the final state of the process of losing something. As suggested by their high covarying collexeme strength (cf. Table 6), some of these expressions appear to represent completely lexically specified instantiations of the pattern: *files go missing*, *chances go begging*.

Instances associated with *go* V-*ing* in Bourdin's 'more grammatical' modal and aspectual senses show in combinations further down the list. They are marginally significant: *you – thinking* (rank 128), *you – blaming* (245), or do not reach significance at all: *bothering* (rank 483), *believing* (678). They express the speaker's disapproval – or warning – and fulfill the constraint postulated for this construction by Bourdin, namely the "mandatory animacy and agentivity of the grammatical subject" (Bourdin 2003: 107).

27. Such uses are subsumed in the OED under sense 44 of the verb *go* "To pass into a certain condition. Chiefly implying deterioration".

Aspectual readings, those of inchoation in (17) and (18) and iteration in (19) – resulting in a proverbial saying – can also be found in the data, again further down the list:

(17) *It may well force* ***you*** *to* ***go looking*** *for more information about your potential audience*. (BNC, AYJ)

(18) ***Chris Fry goes chasing*** *a long ball*. (BNC, K23)

(19) *"He that borrows must pay again with shame or loss"; "****He*** *that* ***goes a-borrowing***, ***goes a-sorrowing***." (BNC, CCT)

The results of the analyses show that go V-*ing* is the most versatile pattern: It comprises several constructions, of which the depictive and purposive stand out. Other readings are more closely associated with lexically specific forms, such as *go missing* and *go begging*, which can be considered as constructions in their own right. The 'more grammatical' instantiations (aspectual and modal readings) get their particulars from the verb *go* and particular contexts respectively, and add to the repertoire of the constructions linked to this pattern.

5. Discussion

5.1. The *go un*-V-*en* construction

The data say that the attributive construction is prominent in the pattern's modern usage. Considering the construction's semantic characterization, the notion of counternormativity identified by Bourdin as its defining feature is not straightforward in some examples, such as (20) and especially (21), which, however, are not central to the pattern. In the collexeme list they take ranks 63 and 36 respectively, but are still attracted to the pattern (highly) significantly (coll. strength: 3.62 and 2.38).

(20) *Now* ***you*** *can* ***go unhampered*** *by disapproval or feelings of guilt and "disloyalty"; and judge whether the activity really is in your line*. (BNC, B3G)

(21) *Instead, he had appeared happy to let* ***her*** *come and* ***go unpestered***, (. . .) (BNC, FP1)

The latter example allows – without considering the wider context in which it was uttered – assuming that *coming* and *going unpestered* is the norm for a woman rather than the opposite. Perhaps Bourdin is aware of

this fact, when he gives "counterexpectation" as a quasi alternative term (cf. above). Counterexpectation, however, is a more general and simultaneously a more subjective and situation-dependent notion than counternormativity: what is counternormative is usually also not expected. Conversely, what is in contrast to one's expectation need not be counternormative, but may be not expected for various other reasons. That is why it needs to be pointed out that the construction's defining feature seems to be counterexpectation rather than counternormativity. This notion follows from the associations found to exist between the covarying subjects and the *un*-V-*en* participles (***warnings*** *going un****heed****ed*, ***needs*** *going un****met*** *needs*, etc). The respective verb-object collocations hint at their status as entrenched units reflecting our expectations. From this perspective, the fact that these events do not take place can indeed be seen as running counter to one's expectations. This is also what may cause the notion of failure (to notice or to communicate) that is found to resonate in many instantiations of the construction (cf. section 2.1), and it may also give rise to describing the depicted scenario as counternormative (Bourdin 2003: 109) or even abnormal (Radden 1996: 449), exactly because it deviates "from the normal and expected course of events".

Complementing Bourdin's semantic specification of the *go un*-V-*en* construction, the collexeme analysis gives specific information about the attracted *un*-participles. The verb groups identified in them hint at the kinds of processes that speakers notice as not taking place (cf. section 4.1.1). That is they are informative about the processes the 'absence' of which is worth communicating, seemingly because this absence goes against expectation. The verb groups can be further conflated, for verbs of mental activity can be seen as metaphorically related to verbs of perception and thus be put in one group. That is, the construction can be said to serve the expression of states resulting from an absence of – predominantly, though not exclusively – mental and communicational processes (of *noticing*, *recognizing*, *reporting*, *mentioning*, for example), and is, hence, not fully productive.

As the attributive was found to originate from the depictive construction, motional readings of *go* were included in our analysis and depictive uses were expected to play an important part in this pattern. Quite unexpectedly, they turned out to be rather scarce: unambiguous depictive examples (such as (22) and (23)) rank rather low in the collexeme list (*unaccompanied* – rank 89, *unblessed* – 46), though they are still (highly) significantly attracted to this *go* pattern (coll. strength 1.52 and 3.01). The co-occurrence with their

subjects also reaches significance: rank 13 and 108 respectively (covarying collexeme strength 2.73 and 2.13).

(22) ***Staff*** *must not* ***go unaccompanied*** *in such situations*. (BNC, GXJ)

(23) *With her grace, and a few prayers the rest of us may find for you,* ***you*** *can hardly* ***go unblessed***. (BNC, G0M)

Moreover, the fact that most of the attracted collexemes can trigger a depictive reading given the appropriate context can be seen as an effect of its past prominence and its ongoing existence.

Along similar lines, it is not surprising to find *unpunished*, the collexeme quoted for the original depictive use, ranking 3rd, as it can occur with both readings, too. The covarying-collexeme analysis reveals that the respective hits instantiate an attributive reading (with inanimate subject referents) rather than a depictive one. Also animate subject referents only rarely trigger depictive readings, while the contextual clues suggest that the persons simply have a lucky escape from a punishable act they were caught in. Interestingly, the covariation of only one of the collexeme pair (of subject and *unpunished*) reaches significance. That means that the other co-occurrences can be attributed to chance and the related scenarios (of *sins*/*mistakes*/*people*/*torture* etc. *going unpunished*) do not seem to be frequent or important enough to become entrenched with a fixed wording. The variability of the subject is, however, not completely unconstrained; on a more schematic level it refers to something that people consider 'punishable'. This restriction holds likewise for the underlying collocations of verb and object and it contributes to the construction's defining feature of counterexpectation, this time with a clear overtone of counternormativity. Besides, animate subject referents (capable of motion) are rare among the top 20 collexemes: there are only three: *staff* going *unaccompanied* (example (22)), *we* going *unbeaten* (in a football season), and *losers* going *uncompensated* (in the economic world). Similar to the last example, animate subject referents from further down the list also trigger an attributive more often than a depictive reading. Thus, judging from the corpus data, *go un*-V-*en* forms are predominantly instances of the attributive construction. That is, Bourdin's strategy to exclude examples of literal *go* proves itself useful and suitable: it – automatically – isolates the attributive from the depictive construction. From a diachronic perspective, the BNC data suggest a change in the prototype of the construction from depictive to attributive uses, which is already alluded to in the OED (cf. sense 23 above). Such a change can be understood as a grammaticalization process of the verb

go from a motion verb to a copular verb and brings about a new construction, which grows in frequency of usage, but does not supersede the depictive use completely. These assumptions need to be corroborated by a diachronic analysis of the construction, which is not pursued here.

5.2. The *go* adjective construction

From the list of attracted collexemes, we had identified three different constructions of the *go* adjective pattern: the depictive (as in *go free|barefoot*), the (subject-related) attributive (as in *go hungry*) and the (subject-related) resultative (as in *go bust|mad|blind*). From these three, the resultative is the predominant construction, being associated with roughly 75% of the top 40 attracted collexemes. The other two are rather marginal, though some of the respective collexemes are significantly attracted, as the attributives *go hungry*, *go strong* (ranks 8, 24), and the depictives *go free* and *go barefoot* (ranks 20, 25) show. Such combinations represent constructions of a lower level of schematicity, that is collocations, some of which also figure in dictionaries.

From the list of significant collexemes, there is one (entrenched) instance not yet specifically referred to: (*go*) *wrong*. The pattern is a bit problematic, since it has two readings: it either exhibits a literal sense, that of taking a wrong way (which is first quoted in the OED from 1300), or it is used – in a metaphorical motion sense – to say that something happens amiss or unsuccessfully (quoted from 1592). So, the pattern as a whole can be understood to express motion, either literal or metaphorical. The latter sense is triggered by inanimate subject referents, such as *things* (incapable of (self-propelled) motion), and the complete phrase simultaneously presents the development of the *things* as gaining a dynamism of its own. Additionally, the meaning of the collexeme *wrong* is somewhat indeterminate. In both the literal and the fictive motion instances, it seems to be associated with an adjunct overtone of direction and/or manner alongside with its expression of a property. Therefore, a conceivable interpretation is to read *wrong* as an adverb expressing direction and/or manner, which is compatible with (and plausible for) the literal-motion sense of *go*. The fictive motion sense, as in *something|things go wrong* (or *fine*), can also be unified with an adjunct reading, the more so since *something* or *things* stand for events rather than objects. This interpretation is supported by the existence of the collocation *things go well*, where the word rendering the sense at issue is explicitly marked as an adverb. Taking the goal bias of *go* into account (see section 2.2), we think that it is not the direction/manner of motion alone that the expression communicates, but it also

alludes to the goal, the final (unsuccessful – or successful) state. That means (*things*) *go wrong* appears to indicate both metaphorical goal-oriented motion and direction, cutting out the source component of the motion schema. The pattern is a classic collocation, and hence a construction in its own right, which allows for little variation: *things go wrong*, *askew*, *everything goes fine*, *everything goes ok/alright*, seemingly on condition that the adjective means roughly the same as *wrong* or *fine*. It is also interesting to note that a negative outcome (*wrong*) is much more prominent (and hence worth communicating) than a positive one. The expression of the latter is less fixed, allowing more variation, which is in line with the negative "aura" of *go* alluded to above.

5.3. The *go* V-*ing* construction

The results of our analyses have shown that the semantic spectrum of *go* V-*ing* is wider than that of the other two patterns. Note also that, again, some V-*ing* complements are involved in more than one construal, depending on the context, and some are even blends. (24), (25) and (26) can serve as examples here:

(24) *Several of his* ***cousins*** *and* ***brothers went flying***. (BNC, CJA) – purposive

(25) *You know when there's like that like a border of virgin snow to stop you, well* ***we went flying*** *into it*. (BNC, KPA) – depictive/manner

(26) *First, make sure you know what you want to do before* ***you go looking*** *for a solution*. (BNC, G00) – inchoation and purpose

All in all, literal-motion readings of *go* are predominant. From the top 40 collexemes, the majority denote a purposive scenario (ranks 1, 2, 4–6, 8, 9 etc.). Collexemes preferably used in depictives are significantly attracted by the construction, too, they, however, rank lower (ranks 7, 10, 23, 29). Attributive instantiations are absent from this pattern. Instead, we have, less centrally, such more grammatical functions associated with the construction as the expression of modal and aspectual notions. The modal connotations of speaker's disapproval or warning (cf. section 4.3.2) have been attributed to the verb *go* as an effect of the diversion schema: diverting from an expected path is not desirable (see also section 4.2.2; Radden 1996: 448; Stefanowitsch 1999: 129, for example). However, taking all analogous examples into consideration, we notice a restriction to non-assertive contexts, more specifically, such expressions are almost always explicitly negated and used as imperatives:

(27) *So **don't** you go telling them any different*. (BNC, ACB)

(28) *I find it very **disturbing** and **unwelcome** to have you go touching me as you do*. (BNC, JYA)

Moreover, we find examples in which the disapproval is superimposed on an otherwise purposive (29) or manner (30) reading:

(29) *"Begging yer pardon, ma'am," Peggy whispered hesitantly, "**you won't go climbing** no steep hills in Corporation Park, will yer?* (BNC, FPK)

(30) *"**You didn't go clambering** over the cliffs, did you?"* (BNC, H8A)

The collexemes involved in such expression (examples (27)–(30)) rank much lower in the collexeme list (ranks 388, 351, 46 and 258), with only *climbing* (more frequent in the purposive construction) reaching high significance (coll.strength 5.83). The others are not significantly attracted and *telling* is even repelled (coll. strength 0.39), though not significantly so. Their co-occurrence with the subject referent *you* reaches significance only once (example (27)), speaking for the existence of the collocation *don't you go telling*. This finding actually supports the idea that the notion of disapproval or warning may be due to the non-assertive context rather than the *go* V-*ing* alone. Collexemes typically inducing an aspectual reading (examples (17)–(19) above) rank quite highly (ranks 13 (coll. strength 2.29), 19 (1.62), 47 (5.81)), but are small in number. The inchoative reading of the pattern may come from the image schema of verb *go*: it implies leaving the deictic centre to move to some new location. This is transferred analogously to beginning or turning to a new action. Iteration does not seem to be motivated by the meaning of *go*, but can rather be read off from the wider context. Instantiations expressing progressing processes with a demoted agent (examples (15) and (16) above) are more prominent (on ranks 3 (coll. strength 184.36) and 16 (20.32)).

6. Conclusion

The results of the analyses presented here allow for a number of conclusions.

Firstly, the data can be interpreted for what they mean for the psychological reality of the patterns identified. The question about the "actual existence", the entrenchment of lower- and higher-level schemas in the minds of the language users is usually operationalized via frequency measures. It is generally assumed (cf. Bybee and Hopper 2001; Bybee 2003,

2006; Diessel 2007, for example) that repetition (and hence frequency) of a linguistic form is crucially involved in its entrenchment. Being concerned with partially schematic and lexically filled patterns, we need to look at both type and token frequencies of the instantiations, because they are assumed to affect entrenchment in different ways. Following Taylor (2002: 277), we assume that "[a] schema [such as *go un*-V-*en*] gains strength (it becomes 'established' or 'entrenched') in proportion to the type frequency of the instances which elaborate it. (...) High token frequency of an instance [such as *things go wrong*] entrenches the instance and weakens (or at least, does not strengthen) the schema."

On the basis of these considerations the type frequencies detected for the instantiations of the three *go* complement$_{\text{Adj}}$ patterns are supportive of their status as constructions (at a mid level of schematicity). The results of the (co-varying) collexeme analyses revealed that each of these three constructions has a prominent function and other more peripheral ones, which sometimes overlap with those of the other constructions. That is, each of the patterns is polysemous, representing a radial category of functionally specific sub-patterns, such as depictive or attributive. These sub-patterns suggest themselves as (sub-)constructions, which, again, is supported by their noticeable type frequencies.

In modern English usage, the go *un*-V-*en* construction is prototypically an attributive construction. Depictives are much less frequent and a peripheral pattern, though some of the clearly depictive instantiations are significantly attracted to the pattern. This is contrary to our expectations that the original function of depictive may be more prominent. What allows for grouping the two patterns into one category is their formal identity (*go* and *un*-V-*en* participles) and their semantic similarity (subject role of AFFECTED and state ascription).

For the *go* adjective construction our expectations are confirmed by the data: the construction can express any function except the purposive. On top of this, our analyses permit a ranking with regard to the centrality of these functions which also form a radial category, with the resultative (sub-)construction in the centre and the depictives and attributives at the periphery. Again, the category as a whole is motivated by similarities in form and meaning. All instantiations contain *go* and an adjective and give information about the state of their subject referents. Depending on the context, the types of subject referents and the types of adjectival collexemes, this state is either construed as attained as the result of change, or it is merely ascribed to a moving or non-moving subject referent (making the whole depictive or attributive). The formal identity of the constructions is indicative of their semantic relationship. The historical records suggest

a similar development as for the *go un*-V-*en* construction: First uses were depictive, from which we get attributives as a result of grammaticalization. Additionally, we have a second development in which literal *go* acquired its resultative meaning, the construction's modern day prototype.

For the *go* V-*ing* construction, the results obtained suggest a more complex category of possible functions than expected. The frequency data indicate that the purposive (sub-)construction is central with the majority of collexemes in the top 40. However, the depictive and ergative uses are also quite strongly represented by significantly attracted collexemes (ranks 7, 10, 29, and 3 and 16). Other collexemes are open to several other readings, which in turn are more peripheral. The fact that some of them co-occur significantly with particular subjects (i.e. with a noticeable token frequency) speaks for them to be more specific instances of the pattern with a construction status of their own (see examples (15) and (16): *files going missing* – rank 15, and *chances going begging* – rank 2).

The latter fact also holds true for a considerable number of instances of the other constructions. That is, apart from the partially schematic patterns, the analyses also revealed patterns that are completely lexically specified as highly attracted collexemes and significantly covarying collexemes are indicative of noticeable token frequencies of such instances (of *go* + collexeme, and subject + *go* + collexeme). These can be assumed to be known by the language user as routines from which the more schematic patterns (for and within the 3 constructional categories) could be abstracted. From this perspective, it is not surprising that most of the collocations correlate with the constructional type found to be prototypical for the respective category, such as the resultatives *go bankrupt* or *go mad* for the *go* adjective construction. A finding like this represents further evidence for the centrality of these constructions.

Secondly, as for the proposed schematic pattern of *go* complement$_{\text{Adj}}$, our frequency data are not informative. Theoretically, this pattern can be thought of as emerging from the three categories of *go* constructions analysed and would in a way "complete" the network of interrelated constructions. This is not implausible, since the extracted data have prominent features in common: they have, as formal components, the verb *go* and a related (quasi-)adjectival element and, as a natural consequence, overlaps have also been found to hold between the extrapolated senses. Such a situation can be understood as a relation of family resemblance. That is, the schematic construction *go* complement$_{\text{Adj}}$ can be seen as a category of polysemous subcategories (*go un*-V-*en*, *go* adjective and *go* V-*ing*). The individual category members cannot be unified on the basis of a common denominator in their semantic pole, they rather form "a criss-crossing

network of similarities" (Taylor 1995: 38), with the *go un*-V-*en* and the *go* adjective patterns sharing the attributive and depictive functions, the *go un*-V-*en* and the *go* V-*ing* patterns sharing the depictive function etc. Moreover (and as already addressed in sections 2.1 and 4.1.2), we also find features shared between the functionally determined (sub-)constructions of depictive, attributive etc. These features are all related to the image-schematic structure associated with the verb *go*. The verb contributes the MOTION schema (SOURCE-PATH-GOAL) and the single constructions profile different components of the schema and blank out others, or draw on more specific MOTION schemas, such as that of DIVERSION. That is, the resulting network can be said to emerge from prototypical (goal-biased) motion, usually rendered by *go* in an intransitive motion construction (X moves Y), with the individual constructions typically highlighting particular aspects of such a motion scenario. In the *go un*-V-*en* construction, depictive and attributive uses highlight (SOURCE and) PATH: X moving in a specified state or persistence of a state through time. The *go* adjective construction is typically resultative with a metaphorical (end-point) GOAL focus, and the *go* V-*ing* construction is mainly purposive with a literal GOAL focus. The relations and metaphorical mappings specified here and more specifically in section 2 seem to reflect two grammaticalization paths the verb *go* undergoes in its historical development. Firstly, the concept of literal motion bleaches via depictives, where motion is backgrounded, to attributives. In the latter uses, *go* can be said to "incorporate" the originally depictive adjectives into its subcategorization frame and expresses the continuation (of the subject referent) in a specified state[28]. Secondly, the resultative may be seen as a separate development where the focus is on the GOAL of motion, giving us the resultant state sense as an effect of the CHANGE IS MOTION and STATES ARE LOCATIONS metaphors. This metaphorisation process also applies to purposive readings and the numerous instantiations of *go wrong*. The purposive reading focuses on the goal component and the semantics of go *wrong* draws on metaphorical motion with an overtone of direction/manner. Admittedly, especially the last considerations need to be tested against diachronic data. We would, however, not have arrived at them without the systematic analysis of such a large data set. The empirical methods employed have shown to be extremely beneficial for the identification of and distinction between constructions which otherwise might have gone unnoticed.

28. Note that also the OED hints at this connection when stating that depictive uses pass into the sense of continuing in a specified state in the early 17th century (first quote from 1610).

Appendix

Table A. Least attracted and repelled covarying-collexeme pairs in the *go un*-V-*en* construction in the BNC

rank	word 1	word 2	freq. w1	freq. w2	obs. w1_2 in_c	exp. w1_2 in_c	relation	cov. coll. strength
98	unnoticed	action	133	3	1	0.74	attraction	0.24
99	unnoticed	behavior	133	3	1	0.74	attraction	0.24
100	unnoticed	development	133	3	1	0.74	attraction	0.24
101	unnoticed	fact	133	3	1	0.74	attraction	0.24
102	unnoticed	piece	133	3	1	0.74	attraction	0.24
103	unnoticed	what	133	3	1	0.74	attraction	0.24
104	unnoticed	activity	133	4	1	0.98	attraction	0.17
105	unnoticed	woman	133	6	1	1.48	repulsion	0.26
106	unnoticed	work	133	6	1	1.48	repulsion	0.26
107	unnoticed	incident	133	5	1	1.23	repulsion	0.19

Table B. Significantly repelled collexemes in the *go* adjective construction in the BNC (coll. strength $>1.3 \rightarrow p < 0.05$)

rank	word	word freq	obs. freq	coll. strength	rank	word	word freq	obs. freq	coll. strength
1	good	75812	5	16.38	17	american	15106	1	3.49
2	different	47607	1	12.86	18	concerned	14663	1	3.37
3	social	41635	1	11.10	19	poor	14562	1	3.34
4	small	41865	2	10.00	20	true	17379	2	3.31
5	british	35305	1	9.25	21	foreign	15950	2	2.94
6	local	43818	5	7.82	22	modern	12888	1	2.89
7	national	37431	3	7.78	23	high	31925	10	2.49
8	long	29616	1	7.60	24	specific	11213	1	2.44
9	full	27228	1	6.92	25	clear	20376	5	2.30
10	great	42710	7	6.18	26	legal	12980	2	2.21
11	special	21662	1	5.33	27	traditional	9714	1	2.04
12	sure	21646	1	5.32	28	ready	8305	1	1.67
13	central	19108	1	4.61	29	low	12909	3	1.66
14	open	19664	2	3.90	30	huge	7648	1	1.50
15	public	16012	1	3.74	31	separate	7580	1	1.49
16	big	24433	4	3.72					

Table C. Significantly repelled collexemes in the *go* V-*ing* construction in the BNC (coll. strength $>1.3 \rightarrow p < 0.05$)

rank	word	word freq	obs. freq	coll. strength	rank	word	word freq	obs. freq	coll. strength
1	going	62663	1	32.75	11	getting	19599	10	3.10
2	making	24643	2	10.73	12	waiting	7447	1	3.02
3	using	24428	3	9.59	13	sitting	7305	1	2.95
4	taking	20828	2	8.78	14	moving	7236	1	2.92
5	saying	16963	2	6.83	15	holding	6233	1	2.43
6	working	16839	2	6.77	16	putting	7267	2	2.23
7	trying	17302	3	6.12	17	seeing	5792	1	2.22
8	following	10366	1	4.48	18	playing	9369	5	1.61
9	giving	12119	3	3.72	19	carrying	5305	2	1.40
10	talking	11877	3	3.62	20	leading	3867	1	1.33

References

Aarts, Bas
1995 Secondary predicates in English. In: Bas Aarts and Charles Meyer (eds.), *The Verb in Contemporary English – Theory and Description*, 75–101. Cambridge etc.: Cambridge University Press.

Auwera, Johan van der and Andrej Malchukov
2005 A semantic map for depictive adjectival. In: Nikolaus Himmelmann and Eva Schulze-Berndt, *Secondary Predication and Adverbial Modification*, 393–421. Oxford: Oxford University Press.

Barlow, Michael and Suzanne Kemmer (eds.)
2000 *Usage-Based Models of Language*. Stanford: Center for the Study of Language and Information Publications.

Bourdin, Philippe
2003 On two distinct uses of *go* as a conjoined marker of evaluative modality. In: Roberta Facchinetti, Manfred Krug & Frank Palmer (eds.), *Modality in Contemporary English*, 103–127. Berlin etc.: Mouton de Gruyter.

Bybee, Joan
2003 Mechanisms of Change in Grammaticization: The Role of Frequency. In: Joseph, Brian D. and Richard D. Janda (eds.), *The Handbook of Historical Linguistics*, 602–623. Malden etc.: Blackwell.

Bybee, Joan
2006 From usage to grammar: The mind's response to repetition. *Language* 82–4: 711–733.

Bybee, Joan and Paul Hopper (eds.)
2001 *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins.
Clark, Eve V.
1974 Normal states and evaluative viewpoints. *Language* 50: 316–331.
Diessel, Holger
2007 Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25: 108–127.
Fillmore, Charles J.
1988 The Mechanisms of 'CG'. In Shelley Axmaker, Annie Jaisser & Helen Singmaster (eds.), *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 35–55. Berkeley: Berkeley Linguistics Society.
Goldberg, Adele
1995 *A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
Goldberg, Adele
2003 Constructions: A new theoretical approach to language. *Trends in Cognitive Science* 7–5: 219–224.
Goldberg, Adele
2006 *Constructions at Work*. Oxford: Oxford University Press.
Goldberg, Adele and Ray Jackendoff
2004 The English Resultative as a family of constructions. *Language* 80(3): 532–568.
Gries, Stefan Th. and Anatol Stefanowitsch
2004 Covarying collexemes in the *into*-causative. In: Michel Achard and Suzanne Kemmer (eds.), *Language, Culture and Mind*, 225–236. Stanford: Center for the Study of Language and Information Publications.
Gries, Stefan Th., Beate Hampe and Doris Schönefeld
2005 Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4): 635–676.
Gries, Stefan Th., Beate Hampe and Doris Schönefeld
2010 Converging evidence II: More on the association of verbs and constructions. In: John Newman and Sally Rice (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*, 59–72. Stanford: Center for the Study of Language and Information Publications.
Halliday, Michael A.K.
1967 Notes in transitivity and theme in English, Part 1. *Journal of Linguistics* 3: 37–81.
Hampe, Beate and Doris Schönefeld
2003 Creative Syntax. Iconic principles within the symbolic. In: Wolfgang G. Müller and Olga Fischer (eds.), *From sign to signing.*

Iconicity in Language and Literature 3: 243–261. Amsterdam: Benjamins.

Hampe, Beate und Doris Schönefeld

2006 Syntactic leaps or lexical variation? – More on "Creative Syntax". In: Stefan Th. Gries and Anatol Stefanowitsch (eds.), *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*, 127–157. Berlin etc.: Mouton de Gruyter.

Himmelmann, Nikolaus and Eva Schulze-Berndt

2005 Issues in the syntax and semantics of participant-oriented adjuncts: an introduction. In: Nikolaus Himmelmann and Eva Schulze-Berndt (eds.), *Secondary Predication and Adverbial Modification*, 1–67. Oxford: Oxford University Press.

Hopper, Paul J.

1987 Emergent grammar. *Berkeley Linguistics Society* 13: 139–57.

Hopper, Paul J. and Elizabeth Closs Traugott

2003 *Grammaticalization.* Cambrigde: Cambrige University Press.

Huddleston, Rodney D. and Geoffrey K. Pullum

2002 *The Cambridge Grammar of the English Language*. Cambridge: CUP.

Hust, Joel R.

1977 The syntax of the unpassive construction in English. *Linguistic Analysis* 3(1): 31–63.

Hust, Joel R.

1978 Lexical redundancy rules and the unpassive construction. *Linguistic Analysis* 4(1): 61–89.

Langacker, Ronald W.

1987 *Foundations of Cognitive Grammar. Volume 1. Theoretical Prerequisites.* Stanford: Stanford University Press.

Langacker, Ronald W.

2000 A Dynamic Usage-Based Model. In Michael Barlow and Suzanne Kemmer (eds.), *Usage-Based Models of Language*, 1–63. Stanford: Center for the Study of Language and Information Publications.

Langacker, Ronald W.

2008 *Cognitive Grammar. A Basic Introduction.* Oxford etc.: Oxford University Press.

Levin, Beth

1993 *English Verb Classes and Alternations*. Chicago etc.: The University of Chicago Press.

Levin, Beth and Malka Rappaport Hovav

1995 *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Cambridge: Massachusetts Institute of Technology Press.

Michaelis, Laura A.

2003 Word meaning, sentence meaning, and syntactic meaning. In: Hubert Cuyckens, René Dirven and John R. Taylor (eds.),

Cognitive Approaches to Lexical Semantics, 163–209. Berlin etc.: Mouton de Gruyter.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik
1985 *A Comprehensive Grammar of the English Language.* London etc.: Longman.

Radden, Günter
1996 Motion metaphorized: The case of *coming* and *going*. In: Eugene H. Casad (ed.) *Cognitive Linguistics in the Redwoods*, 423–458. Berlin etc.: Mouton de Gruyter.

Rapoport, T.R.
1990 Secondary predication and the lexical representation of verbs. *Machine Translation* 5: 31–55.

Rapoport, T.R.
1999 Structure, aspect and the predicate. *Language* 75(4): 653–677.

Rothstein, Susan
2003 Secondary predication and aspectual structure. In: Ewald Lang, Claudia Maienborn and Catherine Fabricius-Hansen (eds.), *Modifying Adjuncts*, 553–590. Berlin etc.: Mouton de Gruyter.

Rothstein, Susan
2006 Secondary predication. In: Martin Everaert and Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, Vol. IV, 209–233. Malden: Blackwell.

Schlüter, Julia
2005 *Rhythmic Grammar.* Berlin etc.: Mouton de Gruyter.

Schönefeld, Doris
2006 Constructions. *Constructions* SV1–1/2006 (www.constructions-online.de, pp. 39).

Stefanowitsch, Anatol
1999 Image schema blending and the construal of events: a cross-linguistic analysis of *go*-and-VERB. In: Dawn Nordquist and Catie Berkenfield (eds.), *Proceedings of the Second Annual High Desert Linguistics Society Conference*, 123–134. Albuquerque: High Desert Linguistics Society.

Stefanowitsch, Anatol and Stefan Th Gries
2003 Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8: 209–243.

Stefanowitsch, Anatol and Stefan Th. Gries
2005 Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1.1: 1–43.

Stefanowitsch, Anatol and Ada Rhode
2004 The goal bias in the encoding of motion events. In: Günter Radden and Klaus-Uwe Panther (eds.), *Studies in Linguistic Motivation*, 249–267. Berlin etc.: Mouton de Gruyter.

Taylor, John R.
1995 *Linguistic Categorization. Prototypes in Linguistic Theory*. Oxford: Clarendon Press.
Taylor, John R.
2002 *Cognitive Grammar*, Oxford etc.: Oxford University Press.
Vendler, Zeno
1957 Verbs and Times. *Psychological Review* 56: 143–160.

The Locative Alternation and the Russian 'empty' prefixes: A case study of the verb gruzit' 'load'*

Svetlana Sokolova, Olga Lyashevskaya, and Laura A. Janda

Abstract

We present an empirical study to address critical aspects of two theoretical issues, namely the "Locative Alternation" and Russian aspectual "empty" prefixes. Our data, extracted from the Russian National Corpus, represent the behavior of the Russian verb *gruzit'* 'load', which participates in the Locative Alternation in both its unprefixed (*gruzit'*) and prefixed forms (*nagruzit'*, *zagruzit'* and *pogruzit'*). According to Russian linguistic tradition, the prefixes *na-*, *za-* and *po-* forming the prefixed counterparts of the verb *gruzit'* 'load' are considered semantically "empty", bearing only the aspectual feature "perfective". The data on the Locative Alternation was analyzed using a logistic regression model in order to probe for a significant relationship between prefixes and grammatical constructions. Our analysis shows that the four verbs behave differently in terms of the locative constructions they participate in (the Theme-Object construction as in *load the hay onto the truck* and the Goal-Object construction as in *load the truck with hay*). While the unprefixed imperfective *gruzit'* favors the Theme-Object construction, the addition of a prefix radically changes this distribution, and each prefix does it in a different way: *nagruzit'* strongly favors the Goal-Object construction, *pogruzit'* uses the Theme-Object construction in a nearly exclusive manner, whereas *zagruzit'* creates a near-balance between the two constructions. Our findings support the hypothesis that the Locative Alternation involves both the meaning of the verb and the meaning of its constructions. The three prefixed verbs exhibit statistically significant differences in their behavior, which is at variance with the idea that the prefixes are semantically empty.

* This research was made possible by a grant from the Norwegian Research Council in support of the "Exploring Emptiness", a research group at the University of Tromsø.

1. Introduction

The present study addresses two theoretical issues, both of which are controversial in the scholarly literature. The first issue is the "Locative Alternation" (*John loaded the hay onto the truck* vs. *John loaded the truck with hay*), where an unresolved debate questions whether the most important factor is a) the meaning of the verb, b) the meaning of the construction, or c) the interaction of both the verb and its construction. Russian provides an excellent testing ground for this issue since we can observe the influence of subtle semantic modifications wrought by prefixes on constructions with overt case marking. The second issue is whether semantically "empty" linguistic units exist. Our data represent the behavior of the Russian verb *gruzit'* 'load', which participates in the Locative Alternation in both its unprefixed and prefixed forms. This verb has three purportedly "empty" prefixes according to traditional definitions, since *nagruzit'*, *zagruzit'* and *pogruzit'* are all listed as the perfective "partners" of the unprefixed imperfective *gruzit'*, and all four verbs come under a single definitional entry (Ožegov and Švedova 2001). Analysis of our data extracted from the Russian National Corpus (www.ruscorpora.ru, henceforth RNC, the source of all examples herein) details the interaction of the verb and construction meanings, supporting hypothesis c) above. Furthermore, since the three prefixed verbs show a significant difference in their distribution of constructions, our data does not support the idea that the prefixes are semantically empty. The rationale is that if the prefixes were semantically empty, they would have to be equivalent, which is not the case. We demonstrate that a verb is not a monolithic unit, since passive participles behave differently from other verb forms. The same "split" applies to the Locative Alternation constructions, which are not uniform and can be represented by their full (see examples 3–5 below) and reduced versions (examples 6–7 below), showing different behavior in terms of reduction. In addition, we find an interesting relationship between the prefixes and the use of prepositions.

Section 2 gives a brief overview of the two theoretical issues, namely the Locative Alternation in 2.1 and the so-called "empty" prefixes in 2.2, situating their relevance to Russian *gruzit'* 'load' in 2.3. Our objectives include probing the relationship between the unprefixed base verb and its three prefixed perfectives and the role of participles and prepositions in *gruzit'* 'load' constructions. Our empirical study presented in Section 3 uses the constructional profile, defined in 3.1 to structure the database, which is described in 3.2. In Section 4, the analysis confronts the objec-

tives with the data, presenting our statistical model in 4.1, addressing the relationship between base and prefixed verbs in 4.2, the behavior of passive participles in 4.3, reduced constructions in 4.4, and prepositions in 4.5. Conclusions are offered in Section 5.0.

2. Theoretical issues

Both the Locative Alternation and the role of prefixes in the Russian aspectual system have produced a vast scholarly literature that we cannot do justice to in this article. Our aim is to survey the highlights of both issues, picking out the points most relevant to our analysis. This entails compressing much of the detail, though this carries some risk of over-simplification.

2.1. The Locative Alternation

The Locative Alternation has been famous in the scholarly literature on English ever since Fillmore (1968: 47) studied examples like these:

(1) Theme-Object: *John loaded the hay onto the truck*

vs.

(2) Goal-Object: *John loaded the truck with hay*.

This phenomenon is observed in many European languages (English, German, Spanish), where a given verb can occur in two alternative constructions, both of which deliver (approximately) the same information. The Locative Alternation has attracted much attention since it touches upon "the fundamental question of why a single verb appears in more than one syntactic frame" (Iwata 2005: 356).

The Locative Alternation has been plagued by terminological diversity. Particularly problematic is the issue of what to call the two constructions, since nearly every author offers a different solution. We choose to follow Brinkmann (1997) and Nichols (2008) in terming the constructions Theme-Object and Goal-Object as above. This pair of terms makes no theoretical assumptions and is relatively transparent. The *hay* item is the theme and the *truck* item is the goal, and "object" refers to the direct object, which is consistently coded with the Accusative case in both constructions in Russian.

Most of the scholarly work on the Locative Alternation can be grouped according to the approach as:

- *Syntactic/lexical* (Rappaport and Levin 1988, 2005, 2008; Pinker 1989, Levin 1993, Brinkmann 1997; Dowty 2000; Mateu 2001);
- *Frame* (Fillmore 1968, 1977, 2008; Boas 2003, 2006); or
- *Constructional* (Goldberg 1995, 2006; Michaelis and Ruppenhofer 2000, 2001; Iwata 2005, 2008).

In a broad sense, all three approaches can be understood as addressing the question of what motivates the Locative Alternation: is it the verb, the construction, or both?

The syntactic/lexical approach focuses on the meaning of the verb as the crucial factor. The syntactic options are viewed as an epiphenomenon of the intrinsic properties of the verb, which can be either "content-oriented" or "container-oriented" (Pinker 1989: 125–127). The option of alternation is listed in the lexicon and follows from linking rules. The goal is thus to determine which sense is basic for each given verb. This approach runs into a variety of problems, among them the claim of derivational direction/complexity (due to the fact that the Goal-Object construction is necessarily more complex in a tree-diagram) and the related claim that one of the verb senses is more basic than the other (see Boas 2006 for an overview and counterexamples). We agree that the meaning of the verb is important, but it does not give discrete results. The syntactic/lexical approach can classify verbs as alternating or non-alternating, but does not account for distributional differences among alternating verbs. We find that alternating verbs can alternate differently, preferring either the Theme-Object or the Goal-Object construction to various extents.

The frame approach takes the syntactic construction rather than the verb as the point of departure. Boas (2006: 135) describes this as a "splitting" approach, where words are defined according to the semantic frames they evoke, and a verb like *load* is split into two lexical units, each of which evokes a distinct frame (the Theme-Object or the Goal-Object construction). Whereas the frame approach highlights the differences between the constructions, it is less effective for investigating why a single verb alternates between constructions.

The Russian data additionally present subtle semantic differencies among *gruzit'* and its three perfectives. All four verbs are glossed as 'load' (Ožegov and Švedova 2001). To some extent, Russian dictionaries regard the prefixed forms under consideration as lexically distinct. All major dictionaries single out two basic "senses" of the unprefixed verb *gruzit'* 'load' 1) 'fill something with freight' and 2) 'place the load somewhere'. Both Ožegov and Švedova (2001) and Evgen'eva (1999) attribute the first meaning to the verb *nagruzit'*, prefixed in *na-*, and the second meaning to the verb

pogruzit' prefixed in *po-*. However, there is no agreement in their judgment of the verb *zagruzit'*: whereas Ožegov and Švedova (2001) group it together with *nagruzit'* as bearing the first meaning, Evgen'eva (1999) does not treat this verb as an aspectual "partner" of the unprefixed verb *gruzit'* at all. In Evgen'eva (1999), *zagruzit'* receives a separate dictionary entry, which in theory contains meanings that characterize this verb as different from other 'load' verbs. However, the first meaning that we find on this list is 'fill something with freight' and the authors do not provide any comments on whether it differs from the meaning of *gruzit'* and *nagruzit'* that is glossed similarly.

The major problem with the traditional lexicographic approach is that dictionaries assume that the distinctions between the 'load' verbs are unilateral: ideally, each meaning of the unprefixed verb should correspond to only one of the prefixed verbs. As we see, this is definitely not the case with *nagruzit'* and *zagruzit'*, which, in fact, overlap not only in the basic meaning 'fill something with freight' but also in the special meaning 'load with work' (Evgen'eva 1999). Furthermore, different dictionaries provide different data: in Ušakov (2009: 704) and Efremova (2006: 772), we find that *pogruzit'* can also be attributed to meaning 1), namely 'fill something with freight'.

Summing up the lexicographic description of the Russian 'load' verbs, we find two kinds of problems. On the one hand, they do not distinguish between constructions and "lexical meanings" (treating both as "lexical meanings"). On the other hand, they usually assign different meanings of the unprefixed verb (defined intuitively) to different prefixed "partners", which in reality is not always the case. A corpus study can provide a more solid ground for distinguishing among the 'load' verbs, showing which factors and in which proportion describe their behavior.

Thus, in the present article, we take corpus data as the point of departure and focus mainly on formal factors and how they are associated with verbal semantics. It appears that the prefixes amplify different portions of the meaning of the base verb and this affects the Locative Alternation. Because we observe this tight interplay between lexical meaning and construction frequency, we choose the constructional approach. We follow Goldberg (1995, 2006) in investigating the dynamics between the Russian 'load' verbs and the Theme-Object vs. Goal-Object constructions. This approach has two added advantages for our analysis. First, the construction approach allows us to examine the interaction between the Locative Alternation constructions and another construction, namely the passive construction. Second, it allows us to zoom in on variation within the Theme-Object construction, targeting the interaction of prefixes and prep-

ositions. Before continuing with this line of argumentation, we need to review the traditional idea of "empty" prefixes in Russian linguistics.

2.2. Russian "empty" prefixes

The category of aspect is consistently expressed by Russian verbs, which can have two values: imperfective or perfective. Janda (2007) demonstrates that it is useful to distinguish among four types of perfective verbs in Russian, two of which are pertinent to this article, namely Natural Perfectives, which serve as the aspectual correlates of imperfective verbs with the same lexical meaning, and Specialized Perfectives that behave as separate lexical items. This distinction can be illustrated with the verb that this study focuses on: *gruzit'* 'load'. *Gruzit' – nagruzit'*, *gruzit' – zagruzit'* and *gruzit' – pogruzit'* form aspectual pairs, where the first member is an imperfective base verb, and the second is its prefixed Natural Perfective (Ožegov and Švedova 2001). Specialized Perfectives like *peregruzit'* 'overload; transport by ship' and *dogruzit'* 'finish loading' involve prefixes that bring new, additional meaning to the imperfective. By contrast, the Natural Perfectives give an impression that their prefix bears no meaning and thus can be treated as "empty".

Specialized perfectives can form their own aspectual correlates by means of the suffixes *-yva-/-iva-*, *-va-* and *-a-* (*peregruzit' – peregružat'* 'overload; transport by ship'). Thus, Russian has two major types of aspectual pairs: 1) unprefixed imperfective verbs and their Natural perfectives, and 2) Specialized perfectives and their suffixal imperfective counterparts. However, this system is further complicated by the fact that many Natural Perfectives can also form suffixal imperfectives, which is also true for the verbs under consideration: *nagruzit' – nagružat'*, *zagruzit' – zagružat'*, *pogruzit' – pogružat'*. Functionally, there is no one-to-one correspondence between primary imperfectives like *gruzit'* and secondary imperfectives like *nagružat'*. The relation between the two types of imperfectives is a separate and complex issue in Russian linguistics and depends on many factors.[1]

1. Secondary imperfectives favor habitual and iterative contexts more than primary imperfectives (see Veyrenc 1980: 166–169; Apresjan 1995: 112–113); in general, secondary imperfectives are more strongly associated with praesens historicum (Petruxina 2000: 99) and are more often used in metaphorical contexts (Veyrenc 1980: 177). Secondary imperfectives reflect not only the interaction of the verbal stem and the perfectivizing prefix, but also involve one more factor, i.e. the imperfectivizing suffix. In this work we are mostly interested in "empty" prefixes, which leaves secondary imperfectives outside the scope of this study.

The idea of "empty" prefixes, also known as "purely aspectual" ("čistovidovyje"), has a long tradition in Russian linguistics (Šaxmatov 1952; Avilova 1959, 1976; Tixonov 1964, 1998; Forsyth 1970; Vinogradov 1972; Švedova 1980; Čertkova 1996; Zaliznjak and Šmelev 2000; Mironova 2004). The list of "pure aspectual" pairs varies in grammars and dictionaries, but, according to the "Exploring Emptiness" database (description of the database is available in Janda and Nesset 2010), there are up to two thousand such pairs used in contemporary Russian. The inventory of "empty" prefixes ranges from sixteen (Švedova et al. 1980) to nineteen items (Krongauz 1998). A noticeable fact about "empty" prefixes is that all these units also form Specialized Perfectives. Usually each base verb chooses one "empty" prefix, but many verbs can occur with two or three prefixes (as in case of *gruzit'*); the maximum appears to be six prefixes (see the description of *mazat'* 'smear' in section 2.3).

Some scholars have objected to the concept of "empty" prefixes, claiming that the prefix always retains its meaning (Vey 1952, van Schooneveld 1958, Isačenko 1960, Timberlake 2004: 410–411). Most traditional descriptions of Russian grammar do not mention the fact that some imperfectives form Natural Perfectives with more than one prefix. Those that do note that Natural Perfectives with various prefixes can be slightly differentiated in lexical meaning (Švedova 1980: 588, Čertkova 1996, Glovinskaja 1982), but do not give further information. We join the camp of opponents of the "meaningless" approach and seek to provide new corpus-based evidence that the prefix of a Natural Perfective has semantic content, and, being compatible with the semantics of the base verb, it enhances or focuses certain portions of the latter.

Janda and Nesset (2010) offer two sets of arguments against the "emptiness" of the prefixes. First we see an uneven distribution of prefixes within the class of Natural Perfectives. If the meanings of the prefixes were really empty, we could expect an arbitrary statistical distribution of verbs to prefixes, which is not the case. Second, there is a remarkable isomorphism between the semantic network of Specialized Perfectives that involve "non-empty" uses of a prefix and the semantic network of Natural Perfectives that use the same prefix in an "empty" mode. This suggests that prefixes always remain connected to their meanings, which overlap with the meanings of the verbs in the Natural Perfectives. The present article provides new evidence against the "empty" prefixes. We demonstrate that the choice of prefix for Natural Perfectives in the case of *gruzit'* (*na-* vs. *za-* vs. *po-*) influences the constructional profile of the verb as it is attested in corpus data.

2.3. Interaction of Locative Alternation and prefixes in Russian

The Locative Alternation is represented by two constructions: Theme-Object and Goal-Object. As noted above, the two constructions differ in which of the participants is marked as the direct object: the theme (i.e. elements like *hay*), or the goal (i.e. elements like *truck*). In both constructions the direct object is consistently coded in Russian with the Accusative case, while the other participant can be expressed via different forms.

The Theme-Object construction encodes the goal via a prepositional phrase (usually with prepositions *v* 'into' and *na* 'onto') with a noun in the Accusative case,[2] as illustrated in examples (3) and (4).

(3) *Potom s pomošč'ju avtokrana predpolagalos' gruzit' brevna na baržu.*
[Then with help-INST crane-GEN was-supposed load-INF logs-ACC on barge-ACC.]
'Then, with the help of the crane, we were supposed to load the logs onto the barge.'

(4) *Gruzi vse v mašinu i vezi sjuda.*
[Load-IMP everything-ACC into car-ACC and bring-IMP here.]
'Load everything into the car and bring [it] over here.'

In the Goal-Object construction the other participant is coded by the Instrumental case without a preposition:

(5) *On sodrogalsja, slušaja o tom, kak gruzili vagony detskimi trupami.*
[He-NOM shuddered hearing about that-LOC how they loaded wagons-ACC childrens'-INST corpses-INST]
'He shuddered hearing about how they loaded wagons with childrens' corpses.'

The use of prefixes in Russian presents a challenge for research on the Locative Alternation in that it introduces a more complicated system of alternating verbs. Considering the interaction between prefixes and locative constructions, three groups of alternating verbs can be singled out:

2. Alternatively adverbs like *kuda* 'in which direction' can appear in this slot of the Theme-Object construction, in which case the goal is not explicitly named.

(a) verbs that can alternate in both unprefixed and prefixed forms (verbs like *gruzit'* 'load');
(b) verbs that do not alternate when unprefixed but are used in both constructions with certain prefixes (verbs like *lit'* 'pour', and *sypat'* 'strew, scatter');
(c) verbs that do not alternate in unprefixed forms and can be used either in Theme-Object or Goal-Object construction depending on the prefix (verbs like *stavit* 'put, place').

The last group is not in our focus since it includes Specialized Perfectives, which are semantically distinct from the imperfective base verb. Hence in this case there is no Locative Alternation as such. For instance, the unprefixed verb *stavit* 'put, place', as well as its Natural Perfective with *po-* (*postavit'*), are used in Theme-Object construction while its Specialized Perfectives with *za-* and *ob-* choose the Goal-Object construction (*zastavit* 'line something with something'; *obstavit* 'furnish').

In group (b) we find Locative Alternation only with a prefix (usually *za-*): cf. the verb *lit'* 'pour', which is used only in the Theme-Object construction, and its Specialized Perfective *zalit'* 'fill', which shows the Locative Alternation (*zalit' benzin*-ACC *v bak*-ACC 'pour gasoline into the tank'; *zalit' bak*-ACC *benzinom*-INST 'fill the tank with gasoline'). It appears that in this case the properties of the prefix are more at stake than the properties of the verbal roots. As well as in group (c), the prefixed verbs of this group are Specialized pefectives and thus go beyond the scope of this article. (For a more detailed consideration of this group see Sokolova and Lewandowski forthcoming.)

Our main interest is in the first group of verbs, which alternate in both unprefixed and prefixed forms. This group is limited in Russian to two sets of verbs: *gruzit'* 'load' and *mazat'* 'smear' and their Natural Perfectives. The verb *gruzit'* has three perfective counterparts, with the prefixes *na-*, *za-*, *po-*, all of which can alternate. The verb *mazat'* 'smear' has six Natural Perfectives, with the prefixes *na-*, *za-*, *po-*, *vy-*, *iz-*, *pro-*, of which only *namazat'* alternates (with a strong preference for the Goal-Object construction).[3] Thus, *gruzit'* 'load' is the only base verb with more or less even dis-

3. It appears that in the case of *mazat'* 'smear' the properties of the verbal root are more at stake than the properties of the prefixes since the verbal root itself already contains some information about the theme as a substance (note the null-suffixed deverbal noun *maz'* 'grease'; cf. verbs with incorporated participants like *saxarit'* 'sugar' derived from *saxar* 'sugar' and *musorit'* 'litter' derived from *musor* 'litter', see Jackendoff 1990; Padučeva 2008: 233–234).

tribution for the Theme-Object and the Goal-Object constructions, where the Natural Perfectives *nagruzit'*, *zagruzit'* and *pogruzit'* can also alternate. Hence it is the behavior of these verbs that we analyze in this article.

3. Data and methodology

Our empirical study examines the constructional profiles of the Russian 'load' verbs as evidenced by data from the Russian National Corpus. We first define the term "constructional profile" and then describe how our data was extracted and coded.

3.1. Constructional profiles

Constructional profiles have proven to be an effective method for investigating the synonymy of words, as Janda and Solovyev (2009: 367) demonstrate in their study of Russian words for 'happiness' and 'sadness', where they define the constructional profile of a word as "the frequency distribution of the constructions that a word appears in". This frequency distribution is based on corpus data.

The constructional profile methodology has grown directly out of the cognitive linguistics tradition, more specifically construction grammar, and has close relatives both within that tradition and beyond it. In keeping with construction grammar, constructional profiling recognizes the construction as the relevant unit of linguistic analysis (Goldberg 1995, 2006) and presumes that speakers are sensitive to the frequency of words in constructions (Goldberg 2006: 46, 62). Both Geeraerts (1988) and Divjak and Gries (Divjak 2006, Divjak and Gries 2006 and Gries and Divjak 2009) have used corpus data to investigate synonymy, using a wide range of factors (collocational, morphosyntactic, syntactic, and semantic) in order to establish behavioral profiles of verbs. Constructional profiles utilize only the complementation patterning aspect of behavioral profiles, specifically targeting the range of constructions a word appears in. Since the constructional profile methodology takes the word as the point of departure, it is in a sense the inverse of the collostructional methodology (Stefanowitsch and Gries 2003, 2005), which takes the construction as the point of the departure and asks what words occur in the construction. Beyond the immediate family of methodologies within cognitive linguistics, constructional profiles are also related to techniques such as syntactic bootstrapping (Gleitman and Gillette 1995, Lidz et al. 2001) and the use of syntactic range information (Atkins et al. 2003).

To a certain extent, our study is parallel to Colleman and Bernolet (this volume). Accepting the claim that the difference between two abstract constructions grants their occurrence with different kinds of verbs, Colleman and Bernolet show that such a split in distribution should be evident not only at the level of *ranges* of verbs that can fill the argument roles of the constructions but also at the level of relative *frequency* with which this occurs. This means that different verbs, as well as different meanings of the same verb, can show different relative frequency distribution across the two constructions.

3.2. Database

According to two dictionaries (Evgen'eva 1999 and Ožegov and Švedova 2001) and a list (Cubberly 1982), the Natural Perfectives of *gruzit'* 'load' include the three prefixed verbs *nagruzit'*, *zagruzit'* and *pogruzit'*. For the purpose of this study, we constructed a database based on the Modern subcorpus (1950–2009) of the RNC, which contains 98 million words. We extracted examples from this subcorpus for each of the four verbs (the base verb and its Natural Perfectives).[4] The same procedure was performed for all verb forms and in addition passive participles received a separate mark.

Passive participles represent an interaction between the Locative Alternation constructions and the passive construction, and this interaction has a significant impact on the distribution of the Locative Alternation constructions. The Locative Alternation involves two objects, Theme and Goal, both of which can be in focus. The passive construction restricts the focus to just one participant. Where non-passive forms show a preference for one construction over the other, this preference is further exaggerated in the presence of passive forms (see Section 4.2). Thus, for the purpose of this study we have treated passive participles as a separate factor. This yields 895 non-passive forms and 1025 passive forms, for a grand total of 1920 examples. Table 1 shows the frequencies of these examples broken down according to verbs.

4. To exclude the author as one more relevant factor, the database was cleaned so that there is only one example for each verb from any single author.

Table 1. Raw frequencies for the forms of the verb *gruzit'* 'load' and its Natural Perfectives

All non-passive forms	raw frequency	Passive participles	raw frequency
gruzit'	286	*gružen*	107
nagruzit'	147	*nagružen*	221
zagruzit'	208	*zagružen*	248
pogruzit'	254	*pogružen*	449

The examples thus accumulated were manually coded for the Locative Alternation constructions as Theme-Object vs. Goal-Object. The breakdown and analysis of these data are presented in 4.2 for the non-passive forms and in 4.3 for the passive forms.

In addition to analyzing the interaction between prefixes and constructions within non-passive and passive forms of the four 'load' verbs, we are also taking into account the subtype of the construction, namely whether the construction is represented by its "full" or "reduced" version. In full constructions, both participants (Theme and Goal) are overtly expressed, while in "reduced" constructions, one of the participants is missing. "Reduction" here refers to the omission of one of the arguments, which is not profiled as a direct object. For the Theme-Object construction this is the case when the Goal is omitted, whereas the Goal-Object construction leaves out the Theme. In most cases with an omitted Theme or Goal argument, the missing participant is perceived from the context, as in examples (6) and (7) given below:

(6) *No uže v bližajšee vremja ožidaetsja podxod sudov obščim tonnažem 780 tys. tonn. Tol'ko zagruzit' ugol' budet problematično, poskol'ku iz-za moroza on prevratilsja v glyby.*

[But already in nearest time is-expected arrival of vessels (Goal that is omitted in the following sentence) with total tonnage 780 thousand tonnes. Just load coal-ACC will-be problematic, since due-to frost-GEN it-NOM turned-into into blocks-ACC.]

'But already very soon we expect the arrival of vessels with total tonnage of 780 thousand tons. Just getting the coal loaded will be problematic since due to the cold it has turned into blocks.'

(7) *Nikolaj . . . očen' skoro upravilsja s pokupkami, nagruzil podvody i, poka mužiki kormili lošadej, otpravilsja slonjat'sja po rjadam.*

[Nikolaj . . . was very soon done with purchases (Theme that is omitted in the following phrase), loaded wagon-ACC and while men were feeding horses he went slouching about rows]

'Nikolaj . . . was very soon done with the purchases, loaded the wagon and while the men were feeding the horses he went slouching about the rows.'

Example (6) illustrates a Theme-Object construction with a missing Goal (the vessels that are mentioned in the previous sentence, where the coal will be loaded), and example (7) illustrates a Goal-Object construction with a missing Theme (the purchases that the wagon is loaded with).[5] Reduced constructions are analyzed in section 4.4.

In the remainder of this article we aggregate data from the full constructions (that name both the theme and the goal) and the reduced constructions.

The reduced constructions frequently involve metaphorical expressions, as in examples (8) and (9), which are parallel to (6) and (7) in structure. Metaphorical uses are a separate and complex issue, and for this reason we do not focus on them in the present article.

(8) *Ja begom kinulsja domoj i, ne razdevajas', vključil komp'juter, zagruzil èlektronnuju kartu goroda.*

[I-NOM run-INST threw-self home and, not having-undressed, turned-on computer-ACC, loaded electronic map-ACC town-GEN.]

'I raced home and turned on my computer without even taking my coat off and downloaded the electronic map of the town.'

(9) *On čto-to vdrug zagruzilsja i rešil zagruzit' svoego predannogo slušatelja.*

[He-NOM somehow suddenly loaded-REFL and decided to-load his-ACC devoted-ACC listener-ACC]

'For some reason he suddenly got confused and decided to confuse his devoted listener.'

5. There were five examples where both the theme and goal were missing, and since in such examples it is not always possible to determine which construction is present, these examples were eliminated from further analysis and do not figure in our database. All five examples involved the unprefixed *gruzit'* 'load'.

Example (8) involves the frame of computer use, where the computer is the CONTAINER, and electronic data are the metaphorical CONTENTS that are loaded into the computer. In example (9), human beings serve as the metaphorical CONTAINERS for information that represents metaphorical CONTENTS. The relationship between metaphorical uses and the reduced constructions is mainly significant for the verb *zagruzit'*, which is further described in Sokolova (forthcoming).

4. Analysis of the Locative Alternation

This study contributes to the ongoing linguistic discussion of what motivates the Locative Alternation by investigating the interaction between the prefixes and the grammatical constructions. First, we look at the relationship between the unprefixed base verb (*gruzit'* 'load') and its prefixed perfective counterparts (*nagruzit'*, *zagruzit'*, *pogruzit'*) to see what the prefixes contribute to the properties of the verbal root. Furthermore, we address an issue which so far has not received proper attention in scholarly works on the Locative Alternation, i.e. the situation with passive participles which change the focus of the locative construction by placing one of the participants (the agent) off-stage. We show that the distribution of the passive participles between the two constructions represents an interaction between the Locative Alternation constructions and the passive construction. Another issue in focus are reduced constructions, where one of the participants is missing. We show that the two constructions behave differently in terms of reduction. Finally, we zoom in on variation within the Theme-Object construction, revealing the interaction of prefixes and prepositions. The data show that the prefix *na-* targets the preposition *na* 'onto' while other prefixes favor the preposition *v* 'into'.

4.1. Binary regression model

The data on the Locative Alternation was analyzed using a logistic regression model in order to probe for a significant relationship between prefixes and grammatical constructions. All calculations were carried out using the "R" software package (http://cran.at.r-project.org), glm, lrm and anova functions (this strategy is modeled after Baayen 2008, Gries 2009[6]).

6. The authors are indebted to an anonymous reviewer for suggesting the use of this method with our data.

Our hypothesis that underlies the model is that three factors, namely 1) prefixes, 2) the number of participants in a frame and 3) the finite/participle form of a verb (as well as their interaction) contribute to the choice of either the Theme-Object or the Goal-Object construction. Thus, there are three independent nominal variables in the model:

1) VERB, having four levels: "Ø" ("zero" for *gruzit'*), "na" (for *nagruzit'*), "za" (for *zaruzit'*) and "po" (for *pogruzit'*);
2) REDUCED, having two levels: "yes" (for the reduced constructions, where one of the participants is missing) and "no";
3) PARTICIPLE, also having two levels: "yes" and "no".

One dependent nominal variable CONSTRUCTION has two levels: "theme" and "goal". The null hypothesis, H_0, suggests that the frequencies of the Theme-Object or the Goal-Object constructions are independent of the VERB, REDUCED, PARTICIPLE variables and their pairwise interactions.

The minimal adequate model retains all the independent variables as main effects, plus the interaction between VERB and PARTICIPLE. As shown below, the unprefixed verb *gruzit'* and its Natural perfective *pogruzit'* favor the Theme-Object construction, while *nagruzit'* and *zagruzit'* prefer the Goal-Object construction. The statistical test also detected that passive participles contribute to the choice of the construction. Finally, reduced frames favor the Goal-Object construction while full frames are used mainly in the Theme-Object construction.

Logistic regression shows that there is a highly significant correlation between the factors mentioned above and the choice of construction: LL-ratio χ^2 (the difference between the two deviance values, with and without predictors) is 1738.47, Nagelkerke's *R2* (correlational strength) is 0.796, *C* value (the coefficient of concordance which according to Gries (2009) should ideally be 0.8 or higher) is 0.964, Somer's D_{xy} (rank correlation between predicted and observed responses) is 0.928, $df = 8$, overall p is 0. The optimal model has high classificatory power: 88.5% constructions are predicted correctly.

The odds ratio, 95%-CI and p for the significant predictors VERB, REDUCED, PARTICIPLE, and VERB:PARTICIPLE are shown in Table 2:

Table 2. Statistical significance of the independent variables and their interactions

Variable	Odds ratio	95%-Confidence Interval		p-value	
VERBna	0.097	5.928746e-02	1.549363e-01	<2e-16	***
VERBpo	79.888	1.744470e+01	1.416632e+03	1.49e-05	***
VERBza	0.289	1.951300e-01	4.245384e-01	3.68e-10	***
REDUCEDyes	0.411	2.907612e-01	5.773928e-01	3.67e-07	***
PARTICIPLEyes	0.003	1.450705e-04	1.203072e-02	4.66e-09	***
VERB na:PARTICIPLEyes	5.881	2.244183e-01	1.541567e+02	0.219043	ns
VERB po:PARTICIPLEyes	289.170	9.203405e+00	9.763774e+03	0.000373	***
VERB za:PARTICIPLEyes	24.057	4.314377e+00	4.521877e+02	0.003034	**

In the next few sections we discuss each factor in more detail.

4.2. The verb *gruzit'* 'load' and its Natural Perfectives

Table 3 shows the distribution of the non-passive forms of *gruzit'* 'load' and its Natural Perfectives across the two constructions of the Locative Alternation. Figure 1 presents the same distribution graphically in terms of relative frequency.

According to our model, the variable VERB has a strong effect ($\chi^2 = 341.52$, $p < 2.2\text{e-}1$). On Figure 1, we see clear differences among the four 'load' verbs. The base imperfective *gruzit'* strongly prefers the Theme-Object construction. The *na-* prefixed perfective is nearly the mirror image, preferring the Goal-Object construction. This preference of *nagruzit'* for

Table 3. Locative Alternation among non-passive forms of *gruzit'* 'load' and its Natural Perfectives

	Theme-Object constructions		Goal-Object constructions		
	raw frequency	relative frequency	raw frequency	relative frequency	Total
gruzit'	208	72.73%	78	27.27%	286
nagruzit'	34	23.13%	113	76.87%	147
zagruzit'	94	45.19%	114	54.81%	208
pogruzit'	253	99.61%	1	0.39%	254

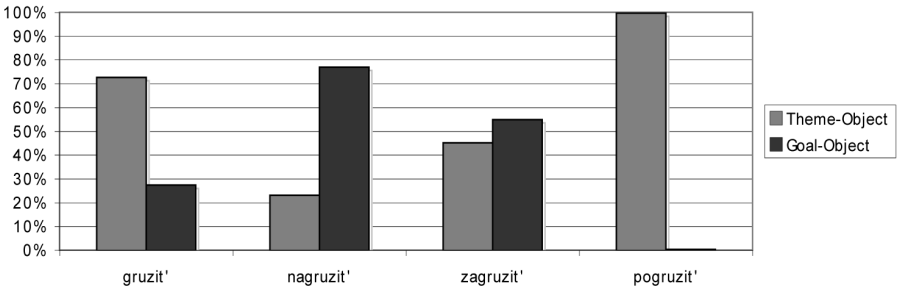


Figure 1. Locative Alternation among non-passive forms of *gruzit'* 'load' and its Natural Perfectives

focusing on the goal may have to do with the SURFACE meaning of *na-*, which corresponds to the meaning of the corresponding preposition *na* 'onto' (which this verb also shows a strong predilection for, see section 4.3). *Zagruzit'* shows an almost even distribution across the two constructions, whereas *pogruzit'* is almost exclusively restricted to the Theme-Object construction, suggesting a focus on the Theme that is loaded rather than the place where the load ends up[7].

Given that the perfective verb *pogruzit'* shows the same focus (i.e. on the Theme) as the unprefixed verb *gruzit'*, *pogruzit'* might seem to be the most natural perfective counterpart of *gruzit'*. However, the fact that the Goal-Object construction constitutes 27% of the total number of uses of *gruzit'* prevents us from making such conclusions. *Pogruzit'* is a natural perfective counterpart of *gruzit'* but only for the Theme-Object construction. Moreover, *gruzit'* and *pogruzit'* behave differently in terms of grammatical forms and reduction (see sections 4.3 and 4.4).

This finding is striking given that all three perfectives are traditionally considered to bear semantically "empty" prefixes. If the three prefixes were indeed empty, we would expect no effect, or at the very least, an identical effect across the three perfectives, i.e. a random distribution. Here, instead,

7. *Zagruzit'* is the only verb that shows an almost even distribution across the two constructions. A more elaborate analysis of the examples indicates that this could be due to the number of additional metaphorical uses that this verb acquires in the Goal-Object construction Of the three prefixed counterparts to the verb *gruzit'* 'load', *zagruzit'* is more often used metaphorically: *zagruzit'* is characterized by 39% of metaphorical uses, while *nagruzit'* and *pogruzit'* have 25% and 11% respectively (see Sokolova and Lewandowski 2010, Sokolova forthcoming).

we find that the three prefixed verbs behave very differently both from the unprefixed imperfective and from each other. We take this as strong evidence against the traditional "empty" prefix hypothesis, since a zero should have no effect, and we cannot countenance three "different" zeroes. As we see below in 4.3, the trends that are evident in the prefixed non-passive forms are even more pronounced in the passive forms.

4.3. Passive participles

Passive participles are used in passive constructions, and here we see an interaction between the two Locative Alternation constructions and the passive construction, as illustrated in examples (10) and (11). The Theme-Object construction has the Theme as the grammatical subject (10), whereas the Goal-Object construction has the Goal as the grammatical subject (11). Whichever item is the grammatical subject is thus strongly profiled, and the agent can be omitted altogether, as we see in both examples.

(10) *K dvum časam vse veščі byli vyneseny na ulicu i pogruženy v avtomobil'.*

[Toward two hours-DAT all things-NOM were carried onto street-ACC and loaded into automobile-ACC.]

'Towards two o'clock all the things were carried out into the street and loaded into the automobile.'

(11) *Pervyj tanker byl zagružen v prisutstvii prezidentov Putina i Nazarbaeva.*

[First tanker-NOM was loaded in presence-LOC presidents Putin and Nazarbaev-GEN.]

'The first tanker was loaded in the presence of presidents Putin and Nazarbaev.'

Table 4 provides the Locative Alternation data for the passive participles of the 'load' verbs. Figure 2 visually presents the same data together with the relative frequencies of non-passive forms for comparison.

Whereas *pogružen* retains its nearly exclusive preference for the Theme-Object construction, all other passive participles have a nearly exclusive preference for the Goal-Object construction. If we look at Figure 2, it may appear that the participles *gružen*, *nagružen*, *zagružen* behave virtually identically. However, they take different objects for the Theme and the Goal and also show different metaphorical representations. For instance, if we compare the metaphorical use of the participles *gružen*, *nagružen*,

Table 4. Locative Alternation among passive forms of *gruzit'* 'load' and its Natural Perfectives

	Theme-Object constructions		Goal-Object constructions		
	raw frequency	relative frequency	raw frequency	relative frequency	Total
gružen	1	0.93%	106	99.07%	107
nagružen	1	0.45%	220	99.55%	221
zagružen	11	4.44%	237	95.56%	248
pogružen	447	99.55%	2	0.45%	449

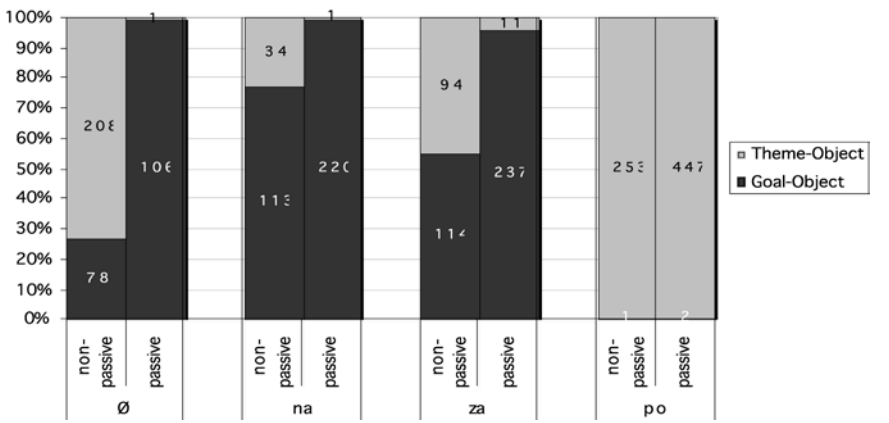


Figure 2. Locative Alternation among non-passive and passive forms of *gruzit'* 'load' and its Natural Perfectives

zagružen we find that *gružen* is hardly ever used metaphorically (2 examples out of 107, about 2%), for *nagružen* metaphorical contexts constitute about 22% (48 out of 221 total), while *zagružen* is characterized by almost 80% metaphorical contexts (176 out of 248).

Not only do participles with different prefixes show different distribution of metaphorical expressions but also the Theme and the Goal in those expressions are represented diffently. One of the most frequent Theme + Goal combinations for *zagružen* is WORK + HUMAN, where the human being serves as a metaphorical CONTAINER for work that represents metaphorical CONTENTS (example (12)):

(12) *Vsju nedelju Ilja byl zagružen delami*
[All week Ilja-NOM was loaded works-INS]
'The whole week Ilja was overloaded with work'

Such contexts exclude the use of *nagružen* (no such examples were attested in the corpus). On the other hand, only the participle *nagružen* can refer to WORDS as a metaphorical CONTAINER and MEANING as their metaphorical CONTENTS (example 13).

(13) *V russkom jazyke nekotorye slova nagruženy negativnym smyslom*
[In Russian language some words-NOM are loaded negative meaning-INS]
'In Russian some words are loaded with negative meaning'

The PARTICIPLE variable demonstrates a significant effect ($\chi^2 = 217.58$, $p < 2.2e\text{-}1$) and at least part of the interaction between VERB and PARTICIPLE (for prefixes *po-* and *za-*) is significant as well ($\chi^2 = 21.5$, $p = 8.284e\text{-}05$, see also Table 2). Our analysis shows that the overall distribution of various constructions within each verb is also dependent on the distribution of grammatical forms within this verb. The frequency of the grammatical form (in our case of the passive participles) is dependent on the verb (for more details see Janda and Lyashevskaya 2011). Some of our verbs show a higher relative frequency of passive participles: for instance, the proportion of non-passive forms to passive forms for the unprefixed verb *gruzit'* is almost 3:1 (286 vs. 107 examples); the verbs *nagruzit'* and *zagruzit'* show an almost even distribution of non-passive and passive forms (1:1.5 and 1:1.2 respectively), while the proportion of the same forms for the verb *pogruzit'* is 1:2 (254 vs. 449 examples).

As can be seen from Figure 2, passive participles have the effect of increasing the relative frequency of the construction that is associated with a given verb. For instance, the distribution of the Theme-Object and Goal-Object constructions with non-passive forms of the verb *nagruzit'* is 23% vs. 77%. For passive forms, the same proportion is 0.5% to 99.5%, significantly increasing the number of examples with the Goal-Object construction. The same effect is attested for the verb *zagruzit'*: the non-passive and passive forms are characterized by a relatively even distribution between the constructions (45% of the Theme-Object constructions vs. 55% of the Goal-Object constructions), while 4.4% passive forms take the Theme-Object constructions and 95.6% take the Goal-Object constructions.

Since passive forms contribute significantly to the overall distribution of the two constructions, the interaction between VERB and PARTICIPLE becomes significant for *pogruzit'* ($p = 0.000373$) and *zagruzit'* ($p = 0.003034$). As a main effect, PARTICIPLE overestimates the probability of the Goal-Object construction because the two other verbs, *gruzit'* and *nagruzit'*, have only one case of the Theme-Object construction with passive forms each. The inclusion of the interaction between VERB and PARTICIPLE more accurately represents this effect in the model.

Thus the passive participles boost the frequency of the construction that is more frequent for non-passive forms. The only exception is the unprefixed verb *gruzit'*, where passive participles change the preference for the construction from the Theme-Object to the Goal-Object. This distribution is the result of general tendencies within the Russian grammatical system, where passive participles are usually formed exclusively from perfective verbs. In those cases where imperfective verbs are characterized by a high frequency of passive participles, they basically perform the function of adjectives: cf. *kopčenyj* 'smoked' as in *kopčenaja ryba* 'smoked fish', *solenyj* 'salted' (*solenye ogurcy* 'pickles', literally 'salted cucumbers'), *žarenyj* 'fried' (*žarenoe mjaso* 'fried meat'). Passive forms of the verb *gruzit'* constitute only ¼ of the data and in the majority of cases characterize the state of the Goal, as in example (14):

(14) *My vozvraščalis'. Navstreču dvigalis' tjaželo gružennye mašiny.*
[We were-going-back. Towards were-moving heavily loaded cars-NOM]
'We were going back. Heavily loaded cars were moving towards us'

In example (14), the participle basically loses its connection with the loading event and mainly refers to the state of the cars, i.e. being heavy.

Thus, the distribution of constructions appears to depend on grammatical forms. Furthermore, as we illustrate in the following section, constructions are sensitive to reduction.

4.4. Reduced constructions

"Reduced constructions" overtly express the participant profiled as the direct object, while omitting the other participant. The tables below provide the frequencies for the reduced structures with non-passive (Table 5) and passive forms (Table 6) of the verb *gruzit'* 'load' and its Natural Perfectives. The same data is made more explicit in Figures 3 and 4.

Table 5. The distribution of reduced structures with non-passive forms of the verb *gruzit'* 'load' and its Natural Perfectives

	Full constructions					Reduced constructions				
All non-passive forms	Theme-Object construction		Goal-Object construction		total	Theme-Object construction		Goal-Object construction		total
	raw fr.	relative fr.	raw fr.	relative fr.		raw fr.	relative fr.	raw fr.	relative fr.	
gruzit'	137	81%	32	19%	169	71	61%	46	39%	117
nagruzit'	27	28%	70	72%	97	7	14%	43	86%	50
zagruzit'	64	51%	62	49%	126	30	37%	52	63%	82
pogruzit'[8]	207	100%	0	0%	207	46	98%	1	2%	47

Table 6. The distribution of reduced structures with passive forms of the verb *gruzit'* 'load' and its Natural Perfectives

	Full constructions					Reduced constructions				
Passive forms	Theme-Object construction		Goal-Object construction		total	Theme-Object construction		Goal-Object construction		total
	raw fr.	relative fr.	raw fr.	relative fr.		raw fr.	relative fr.	raw fr.	relative fr.	
gružen	1	1%	90	99%	91	0	0%	16	100%	16
nagružen	1	0.7%	134	99.3%	135	0	0%	86	100%	86
zagružen	6	6%	95	94%	101	5	3.4%	142	96.6%	147
pogružen	427	100%	0	0%	427	20	91%	2	9%	22

8. The diagram does not include the verb *pogruzit'* since it is almost never attested in the Goal-Object construction and the interaction between reduction and the construction does not seem to be relevant.

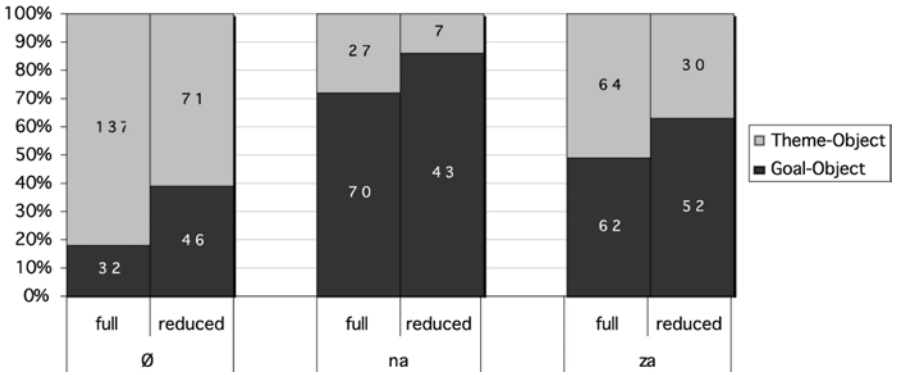


Figure 3. The distribution of reduced structures with non-passive forms of the verb *gruzit'* 'load' and its Natural Perfectives

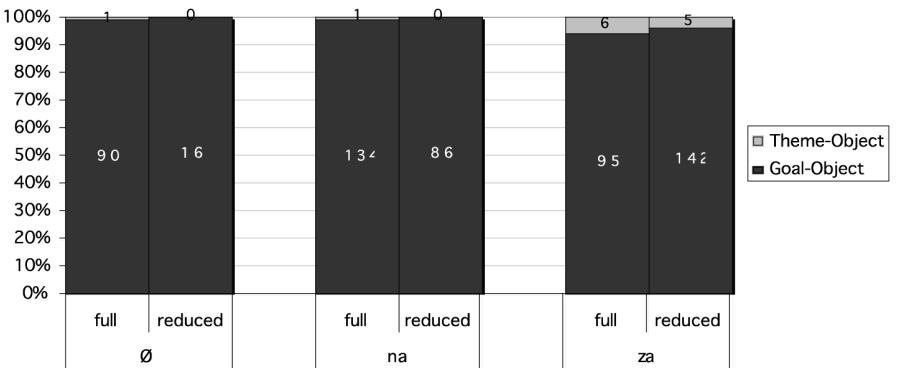


Figure 4. The distribution of reduced structures with passive forms of the verb *gruzit'* 'load' and its Natural Perfectives

The REDUCED variable has a significant correlation with the choice of the construction ($\chi^2 = 26.8$, $p = 2.257\text{e-}07$). As can be seen from Figure 3, the Goal-Object construction shows a higher frequency with reduced constructions: about 20% higher for *gruzit'* and *nagruzit'* and 14% higher for *zagruzit'*. This proportion illustrates that the two constructions behave differently in terms of reduction. Furthermore, the only contexts where the verb *pogruzit'* is attested in the Goal-Object construction are reduced structures, as illustrated by example (15)):

(15) *… mašinu uže pogruzili … značit ona … s instrumentom/ da?*
[Car-ACC already they-loaded … so it-NOM … with tools-INS/ yes?]
'The car has already been loaded … So, the tools are already there, right?'

The car, represented as a direct object, is the Goal in the construction since the following context specifies that the car contains the tools, which are the Theme.

One more important difference between the Theme-Object and the Goal-Object constructions in terms of their relation to reduction is that the quality of reduced structures in the two constructions appears to be different. In examples (6) and (7), the missing component is mentioned in the previous context and thus can be treated as an instance of ellipsis. Such cases are attested for both the Theme-Object and the Goal-Object construction. Yet, the Goal-Object construction is also characterized by cases where reduction interacts with metaphor. The major metaphorical extensions involve a "person" (Goal), who serves as the metaphorical CONTAINER, and "information" or "work" (Theme), which represents metaphorical CONTENTS, as shown in example (9) above and examples (16)–(17) below:

(16) *A ty, Volodin, nas togda nagruzil pro vnutrennego prokurora.*
[And you-NOM, Volodin-NOM, us-ACC then loaded about internal prosecutor-ACC.]
'And you, Volodin, completely confused us then concerning the internal prosecutor.'

(17) *Koroče, on nagruzil artistov tak, čto v itoge my snjali xorošee kino.*
[In-short, he-NOM loaded artists-ACC so, that in end we shot good-ACC film-ACC]
'In short, he stressed the artists so much that we ended up shooting a good film.'

In example (16), a human being (the listener) serves as the metaphorical CONTAINER for information that represents metaphorical CONTENTS. Analogously, in (17), the human beings (the artists) are loaded with work. Such contexts should be distinguished from cases of ellipsis since the omission of the second participant is highly conventionalized. In Fillmore's terminology, sentences like (16) and (17) can be treated as "definite null instantiations" of the Theme, when a participant is consistently omitted and is

not mentioned in the preceding context, but is known to the speaker and the hearer (Fillmore 2008).

The Theme-Object constructions can also involve both metaphor and reduction, but such structures are less frequent than the Goal-Object construction and the missing component is usually present in the previous context (see example (8)):

(8) *Ja begom kinulsja domoj i, ne razdevajas', vključil kompjuter* (the Goal that is further omitted)*, zagruzil èlektronnuju kartu goroda.*

[I-NOM run-INST threw-self home and, not having-undressed, turned-on computer-ACC, loaded electronic map-ACC town-GEN.]

'I raced home and turned on my computer without even taking my coat off and downloaded the electronic map of the town.'

In addition to the three correlations discussed above (between the construction and such factors as the verb, the grammatical form and reduction), our data also shows a correlation between the prefix and prepositions. This correlation can be attested only in the full version of the Theme-Object construction, for which reason we did not include it in our regression model. The next subsection examines the role of prepositions in more detail.

4.5. Prepositions

As discussed above, the non-passive forms of *nagruzit'* strongly prefer the Goal-Object construction, and there might be a connection here between the SURFACE meaning of the prefix *na-* and its etymological cousin, the preposition *na* 'onto'. The focus on surfaces suggests a focus on locations (goals) as opposed to goods (themes) that are loaded on them. Because prepositions are used only in the Theme-Object construction, all data in this subsection pertains only to that construction.

Table 7 shows the distribution of prepositions that occur in the Theme-Object construction. The right-most column in Table 7, marked "no preposition", aggregates a variety of types of data, since the path of the Theme can alternatively be marked by various adverbs or omitted altogether. Figure 5 presents the same data in terms of percentages (ignoring the uses without a preposition) graphically.

In order to probe for a significant relationship between prefixes and prepositions, the data in Table 7 was analyzed using χ^2-test, excluding the "no preposition" column, which is heterogeneous and thus not strictly comparable to the data in the other two columns. A χ^2-test comparing the

Table 7. Prepositions used with non-passive forms of 'load' verbs to mark the goal in the Theme-Object construction

	preposition *na* 'onto'	preposition *v* 'into'	no preposition
gruzit'	67	67	66
nagruzit'	19	2	3
zagruzit'	7	52	35
pogruzit'	54	143	55

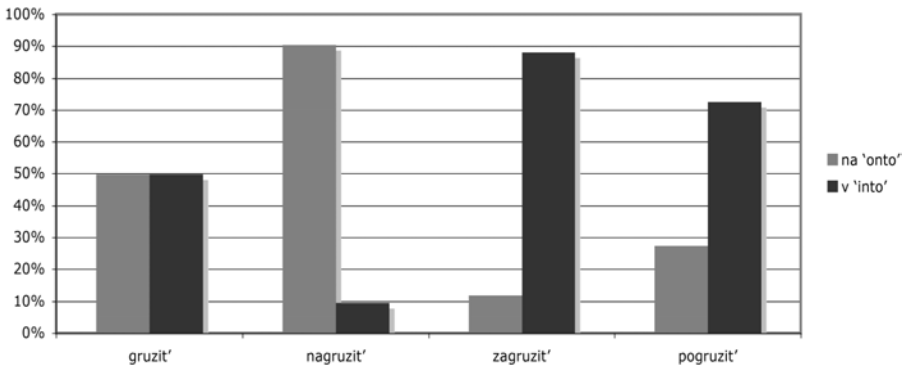


Figure 5. Prepositions used with non-passive forms of 'load' verbs to mark the goal in the Theme-Object construction

distribution of frequencies yields a value of 59.8343 (df = 3, $p = 6.377e\text{-}13$), suggesting an association between the choice of the prefix and the choice of the preposition. To measure the effect size of the χ^2 values, Cramer's V was used, where 0.1 is a small size, 0.3 is a moderate size, and 0.5 is a large size (Cohen 1998: 215–271; King and Minium 2008: 327–330). In our case, the effect size measured by Cramer's V is 0.38, thus registering between a moderate and a large effect.

The imperfective base verb *gruzit'* 'load' has no preference with regard to the prepositions *na* 'onto' and *v* 'into'. *Nagruzit'* attracts the preposition *na* 'onto', while both *zagruzit'* and *pogruzit'* follow the opposite trend, attracting the preposition *v* 'into'. It appears that the choice of the preposition in the Theme-Object construction depends on whether the goal is understood as a SURFACE (*na* 'onto') or as a CONTAINER (*v* 'into'). The association of the *na-* prefixed verb with the preposition *na* makes sense,

since the preposition and the prefix have inherited a meaning that refers to a SURFACE, cf. the verb *nadet'* 'put on (clothing)' and the phrase *na stol* 'onto the table'. This connection is palpable also in examples like (17):

(17) *Na teležku nagruzili celuju goru jaščikov, čemodanov i meškov.*

[Onto cart-ACC loaded whole mountain-ACC boxes, suitcases and bags-GEN.]

'[They] loaded a whole mountain of boxes, suitcases and bags onto the cart.'

Zagruzit' and *pogruzit'*, on the other hand, strongly prefer the preposition *v* 'into', where the goal is conceptualized as a CONTAINER, as in (18) and (19).[9]

(18) *Krome togo, v mašinu zagruzili ogromnyj rjukzak s paraplanom, paru kanistr, veščі, instrument i koe-kakuju meloč'.*

[Beside that-GEN, into car-ACC loaded huge backpack-ACC with paraglider-INST, pair-ACC canister-GEN, things-ACC, instrument-ACC and various trifles-ACC.]

'In addition [they] loaded a huge backpack with a paraglider, a couple of canisters, things, an instrument and various trifles into the car.'

(19) *Pogruziv s pomošč'ju šofera v mašinu svoi veščі, Tamara vsju dorogu do goroda prodremala.*

[Having-loaded with help-INST driver-GEN into car-ACC own things-ACC, Tamara-NOM whole way-ACC to town-GEN slept.]

'Having loaded her things into the car with the driver's help, Tamara slept all the way to town.'

9. In the case of *zagruzit'*, this preference may be due to a parallelism between the preposition *v* 'into' and the preposition *za* 'beyond', both of which can refer to crossing the boundary of a container. In the case of *pogruzit'*, the preference for *v* 'into' may be explained by the presence of some examples that continue the original meaning of this verb as 'sink, plunge', from which the 'load' meaning is historically derived via metonymy (since barges sink when loaded, cf. Nichols 2008). These are, however, speculative remarks that will need further study.

5. Conclusions

The constructional profiles of the four Russian 'load' verbs, *gruzit'*, *nagruzit'*, *zagruzit'*, and *pogruzit'* are distinct: logistic regression shows that there is a highly significant correlation between the verb and the choice of the construction. This finding supports the theoretical hypothesis that the meanings of words and constructions interact, as suggested by the constructional approach to the Locative Alternation. The syntactic/lexical-semantic approach cannot account for the observed variation among verbs, since it can only recognize verbs as having the alternation or lacking it. The frame approach would constrain us to treating each of the 'load' verbs as a pair of homonyms, and again we would lose sight of the differences in variation.

The unprefixed imperfective *gruzit'* favors the Theme-Object construction. The addition of a prefix radically changes this distribution, each in a different way: *nagruzit'* strongly favors the Goal-Object construction, *zagruzit'* creates a near-balance between the two constructions, whereas *pogruzit'* uses the Theme-Object construction in a nearly exclusive manner. This finding contradicts the traditional assumption that the prefixes *na-*, *za-*, and *po-* function as semantic zeroes in forming perfective partner verbs from *gruzit'*. If the prefixes were zeroes, they should follow a random distribution (since they all perfectivize the verb).

The observation of three distinct effects indicates that the prefixes are not devoid of meaning. There is, however, a way to reconcile this finding with the traditional understanding of "purely aspectual" prefixes if we recognize the effect of the prefixes as semantic overlap rather than merely addition. Because the meanings of the prefixes and the verb overlap, there is an illusion of emptiness (cf. Janda and Nesset 2010). Our data show that even these overlaps result in dramatic differences in the constructional profiles of the resulting perfectives.

Furthermore, there appears to be an interaction between the two Locative Alternation constructions and the passive construction. The past passive participles largely suppress the Locative Alternation, using the Goal-Object construction, except in the case of *pogruzit'*, where the nearly exclusive preference for the Theme-Object construction remains. A possible explanation of this distribution is that passive participles boost the frequency of the main construction associated with the verb (Goal-Object for *nagruzit'* and *zagruzit'*, and Theme-Object for *pogruzit'*), perhaps due to the focus of attention on the patient. The unprefixed verb *gruzit'*, where

passive participles change the preference from the Theme-Object to the Goal-Object construction, appears to be an exception caused by the general tendencies within the Russian grammatical system. In Russian, passive participles are formed primarily from perfective verbs. When formed from imperfective verbs, participles usually perform the function of adjectives, which in the case of *gruzit'* characterize the state of the Goal. This finding requires further investigation on a larger number of verbs.

Both Theme-Object and Goal-Object constructions can be represented via reduced versions, where the former omits the Goal and the latter omits the Theme. Our model also shows that there is a correlation between the construction and its full or reduced version: reduced frames favor the Goal-Object construction, while full frames are used mainly in the Theme-Object construction. The interaction between the Goal-Object construction and reduction is supported by two observations: on the one hand, the Goal-Object construction shows a higher frequency with reduced constructions for the verbs *gruzit'*, *nagruzit'* and *zagruzit'*; on the other hand, reduced structures are the only contexts where the verb *pogruzit'* is attested in the Goal-Object construction. One more important difference between the Theme-Object and the Goal-Object constructions in terms of their relation to reduction is that the quality of reduced structures in the two constructions appears to be different: in the case of the Theme-Object construction, we mostly deal with ellipsis, where the missing component is mentioned in the previous context, while the Goal-Object construction is also characterized by conventionalized reduced contexts, where reduction interacts with metaphor. The major metaphorical extensions here involve a "person" (Goal), who serves as the metaphorical CONTAINER, and "information" or "duties" (Theme), which represent metaphorical CONTENTS. This topic merits further research.

Within the Theme-Object construction, we find an interesting distribution of prepositions. Whereas the unprefixed imperfective *gruzit'* shows a three-way split among use of the preposition *na* 'onto', *v* 'into' and no preposition, the prefixed perfectives have strong preferences. The prefix *na-* in *nagruzit'* prefers its etymological cousin *na* 'onto', but both *za-* and *po-* prefer *v* 'into'. It may be that *nagruzit'* is primarily used with goals that are understood as surfaces, whereas *zagruzit'* and *pogruzit'* tend to select for goals that are understood as containers. However, there is considerable variation here and this topic can also be taken up in future work.

References

Apresjan, Jurij D.
1995 Traktovka izbytočnyx aspektual'nyx paradigm v tolkovom slovare [Rendering superfluous aspectual paradigms in an explanatory dictionary]. *Izbrannye trudy*, Volume 2 (Integral'noe opisanie jazyka [Integral description of language]), 102–113.

Atkins, Sue, Charles J. Fillmore and Christopher R. Johnson
2003 Lexicographic relevance: Selecting information from corpus evidence. *International Journal of Lexicography* 16: 251–280.

Avilova, Natal'ja S.
1959 O kategorii vida v sovremennom russkom literaturnom jazyke [On the category of aspect in modern Russian]. *Russkij jazyk v nacional'noj škole* 4: 21–26.

Avilova, Natal'ja S.
1976 *Vid glagola i semantika glagol'nogo slova* [Verbal aspect and the semantics of verbal units]. Moscow: Nauka.

Baayen, R. Harald
2008 *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge UP: Cambridge.

Boas, Hans C.
2003 "A Lexical-constructional Account of the Locative Alternation," in: Carmichael, L., C.-H. Huang, and V. Samiian (eds.), *Proceedings of the 2001 Western Conference in Linguistics* 13: 27–42.

Boas, Hans C.
2006 A frame-semantic approach to identifying syntactically relevant elements of meaning. In: Steiner, P., Boas H.C., and S. Schierholz (eds.), *Contrastive Studies and Valency. Studies in Honor of Hans Ulrich Boas*, 119–149. Frankfurt/New York.

Brinkmann, Ursula
1997 *The Locative Alternation in German. Its structure and acquisition*. Amsterdam/Philadelphia: John Benjamins.

Čertkova, Marina Ju.
1996 *Grammatičeskaja kategorija vida v sovremennom russkom jazyke* [The grammatical category of aspect in modern Russian]. Moscow: Moscow State University.

Cohen, Jacob
1988 *Statistical Power Analysis for the Behavioral Sciences*. Mahwah, New Jersey/London: Lawrence Earlbaum Associates.

Colleman, Timothy and Sarah Bernolet
This volume Divjak, Dagmar, and Stefan Gries (eds.), *Frequency effects in linguistic representations*. Berlin: Mouton de Gruyter. Trends in Linguistics Series.

Cubberly, Paul V.
1982 On the 'empty' prefixes in Russian. *Russian Language Journal* 36: 14–30.
Divjak, Dagmar
2006 Ways of intending: Delineating and structuring near-synonyms. In: Gries, Stefan, and Anatol Stefanowitsch (eds.), *Corpora in Cognitive Linguistics*, Volume 2: *The Syntax-Lexis Interface*, 19–56. Berlin: Mouton de Gruyter.
Divjak, Dagmar and Stefan Th. Gries
2006 Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2: 23–60.
Dowty, David
2000 The Fallacy of 'Argument Alternation'. In: Ravin, Y., and C. Laecock (eds.), *Polysemy*, 111–128. Oxford: Oxford University Press.
Efremova, Tat'jana F.
2006 *Sovremennyj tolkovyj slovar' russkogo jazyka* [Modern Explanatory Dictionary of Russian], Volume 2. Moscow: AST: Astrel'.
Evgen'eva, Anastasija P. (ed)
1999 *Slovar' russkogo jazyka* [Dictionary of Russian]: In four volumes, Moscow: Russkij jazyk.
Fillmore, Charles J.
1968 The Case for Case. In: E. Bach and R. T. Harms (eds.), *Universals in Linguistic Theory*, 1–88. New York: Holt, Rinehart and Winston.
Fillmore, Charles J.
1977 The Case for Case Reopened. In P. Cole (ed.), *Syntax and Semantics 8: Grammatical Relations*, 59–81. New York Academic Press.
Fillmore, Charles J.
2008 Border Conflicts: FrameNet Meets Construction Grammar (Presented as a plenary talk at *Euralex* 2008), Manuscript, University of California, Berkeley.
Forsyth, James A.
1970 *Grammar of Aspect*. Cambridge: Cambridge University Press.
Geeraerts, Dirk
1988 Where does prototypicality come from? In: Rudzka-Ostyn, Brygida (ed.), *Topics in Cognitive Linguistics*, 207–229. Amsterdam: John Benjamins.
Gleitman, Lila R. and Jane Gillette
1995 The role of syntax in verb learning. In: Fletcher, Paul and Brian MacWhinney (eds.), *The Handbook of Child Language*, 413–427. Oxford: Blackwell.
Glovinskaja, Marina Ja.
1982 *Semantičeskie tipy vidovyx protivopostavlenij russkogo glagola* [Semantic types of aspectual oppositions of the Russian verb]. Moscow: Nauka.

Goldberg, Adele E.
1995 *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago/London: The University of Chicago Press.

Goldberg, Adele E.
2006 *Constructions at Work: The Nature of Generalizations in Language*. Oxford: Oxford University Press.

Gries, Stefan Th.
2009 *Statistics for Linguistics with R: A Practical Introduction*. Berlin/New York: Mouton de Gruyter.

Gries, Stefan Th. and Dagmar Divjak
2009 Behavioral profiles: a corpus-based approach towards cognitive semantic analysis. In: Evans, Vyvyan and Stephanie S. Pourcel (eds.), *New directions in cognitive linguistics*, 57–75. Amsterdam and Philadelphia: John Benjamins.

Isačenko, Aleksandr V.
1960 *Grammatičeskij stroj russkogo jazyka v sopostavlenii s slovackim. Morfologija* [Russian grammatical system in comparison with Slovak. Morphology]. Bratislava: Izdatel'stvo akademii nauk.

Iwata, Seizi
2005 Locative Alternation and the two levels of verb meaning. *Cognitive Linguistics* 16(2): 355–407.

Iwata, Seizi
2008 *Locative Alternation: A lexical-constructional approach*. Amsterdam/Philadelphia: John Benjamins.

Jackendoff, Ray S.
1990 *Semantic Structures*. Cambridge: MIT Press.

Janda, Laura A.
2007 Aspectual Clusters of Russian Verbs. *Studies in Language* 31(3): 607–648.

Janda, Laura A. and Tore Nesset
2010 Taking Apart Russian RAZ. *Slavic and East European Journal* 54(3): 476–501.

Janda, Laura A. and Olga Lyashevskaya
2011 Grammatical profiles and the interaction of the lexicon with aspect, tense and mood in Russian. *Cognitive Linguistics* 22(4): 719–763.

Janda, Laura A. and Valery Solovyev
2009 What Constructional Profiles Reveal About Synonymy: A Case Study of Russian Words for SADNESS and HAPPINESS. *Cognitive Linguistics* 20: 367–393.

King, Bruce M. and Edward Minium
2008 *Statistical Reasoning in the Behavioral Sciences*. Hoboken, NJ: John Wiley and Sons.

Krongauz, Maksim A.
1998 *Pristavki i glagoly v russkom jazyke: semantičeskaja grammatika* [Prefixes and verbs in Russian: A semantic grammar]. Moscow: Jazyki russkoj kul'tury.

Levin, Beth
1993 *English verb classes and alternations: a preliminary investigation*. Chicago: University of Chicago Press.

Lidz, Jeffrey, Henry Gleitman and Lila Gleitman
2004 Kidz in the 'hood: Syntactic Bootstrapping and the Mental Lexicon. In: Hall, D. G. and S. R. Waxman (eds.), *Weaving a Lexicon*, 603–636. Cambridge, MA: MIT Press.

Mateu, Jaume
2001 *Lexicalization patterns and the Locative Alternation*. GGT-01-5 Research Report. Universitat Autònoma de Barcelona, Bellaterra.

Michaelis, Laura A. and Josef Ruppenhofer
2000 Valence Creation and the German Applicative: the Inherent Semantics of Linking Patterns. *Journal of Semantics* 17(4): 335–395.

Michaelis, Laura A. and Josef Ruppenhofer
2001 Beyond Alternations: A constructional model of the German applicative pattern. Stanford: CASLI Publications. University of Chicago Press.

Mironova, L. Ju.
2004 *Vid glagola i sootnošenie leksiko-semantičeskix variantov mnogoznačnogo slova* [Verbal aspect and the correlation between lexical-semantic variants of a polysemous word]. Tambov: Avtoreferat.

Nichols, Johanna
2008 Prefixation and the Locative Alternation in Russian contact verbs. Presentation at the annual conference of the American Association of Teachers of Slavic and East European Languages in San Francisco.

Ožegov, Sergej I. and Natalija Ju. Švedova
2001 *Slovar' russkogo jazyka* [The Russian Language Dictionary]. Moscow: Russkij jazyk.

Padučeva, Elena V.
2008 The Semantic-Syntactic Interface: Incorporated Actants and Excorporation. *Automatic Documentation and Mathematical Linguistics* 42(4): 232–238. Allertion Press, Inc.

Petruxina, Elena V.
2000 *Aspektual'nye kategorii glagola v russkom jazyke v sopostavlenii s češskim, slovackim, pol'skim i bolgarskim jazykami* [Russian aspectual verbal categories in comparison with Czech, Slovak, Polish and Bulgarian]. Moscow: Moscow State University.

Pinker, Steven
1989 *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.

Rappaport Hovav, Malka and Beth Levin
1988 What to do with theta-roles. In: Wilkins, W. (ed.), *Syntax and Semantics* 21: 7–36. (Thematic Relations). San Diego, CA: Academic Press.
Rappaport Hovav, M. and Beth Levin
2005 Change of State Verbs: Implications for Theories of Argument Projection. In: N. Erteschik-Shir and T. Rapoport (eds.), *The Syntax of Aspect*, 274–286. Oxford: Oxford University Press.
Rappaport Hovav, M. and Beth Levin
2008 Reflections on Manner/Result Complementarity. Unpublished manuscript, Stanford University and the Hebrew University of Jerusalem.
Šaxmatov, Aleksej A.
1952 *Učenie o častjax reči* [A theory of parts of speech]. Moscow: Učebno-pedagogičeskoe izdatel'stvo.
Švedova, Natalija Ju. (ed.)
1980 *Russkaja grammatika* [Russian Grammar], Volume 1. Moscow: Nauka.
Schooneveld, Cornelius H. van
1958 The so-called 'préverbes vides' and neutralization. *Dutch contributions to the Fourth International Congress of Slavistics*, 159–161. The Hague: Mouton.
Sokolova, Svetlana and Wojciech Lewandowski
2010 Constructional Profile of the Verbal Prefix ZA-: A Comparative Study of Russian and Polish. In A. Grønn & I. Marijanovic (eds.): Russian in Contrast, *Oslo Studies in Language* 2(2): 365–391.
Sokolova, Svetlana
Forthcoming *Asymmetries in Linguistic Construal: Russian Prefixes and the Locative Alternation*. Ph.D. dissertation. The Faculty of Humanities, Social Sciences and Education, University of Tromsø.
Stefanowitsch, Anatol and Stefan Th. Gries
2003 Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8: 209–243.
Stefanowitsch, Anatol and Stefan Th. Gries
2005 Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1: 1–43.
Timberlake, Alan
2004 *A Reference grammar of Russian*. Cambridge: Cambridge University Press.
Tixonov, Aleksandr N.
1964 Čistovidovye pristavki v sisteme russkogo vidovogo formoobrazovanija [Purely aspectual prefixes in the system of Russian verbal word-formation]. *Voprosy jazykoznanija* 1: 42–52.
Tixonov, Aleksandr N.
1998 *Russkij glagol* [The Russian Verb]. Moscow: Russkij jazyk.

Ušakov, Dmitij N.
2009 *Bol'šoj tolkovyj slovar' sovremennogo russkogo jazyka* [Full Explanatory Dictionary of Modern Russian]. Moscow: Al'ta-Print, Dom XXI vek.

Vey, Marc
1952 Les préverbes 'vides' en tchéque moderne. *Revue des études slaves* 29: 82–107.

Veyrenc, Jacques
1980 *Études sur les verbe russe*. Paris. Institut d'études slaves.

Vinogradov, Viktor V.
1972 *Russkij jazyk* [The Russian Language]. Moscow: Vysšaja škola.

Zaliznjak, Anna. A. and Aleksej D. Šmelev
2000 *Vvedenie v russkuju aspektologiju* [An introduction to Russian aspectology]. Moscow: Jazyki russkoj kul'tury.

Alternation biases in corpora vs. picture description experiments: DO-biased and PD-biased verbs in the Dutch dative alternation

Timothy Colleman and Sarah Bernolet[1]

Abstract

In semantic studies of argument structure alternations as well as in recent psycholinguistic research on syntactic priming, the concept of *alternation bias*, i.e. the lexical preferences of individual verbs for one of two (or more) alternating constructions, plays a crucial role. This paper offers a detailed comparison of the results from Colleman's (2009) corpus-based investigation of the dative alternation in Dutch with the findings from a series of picture description experiments reported in Bernolet (2008). On the one hand, this comparison reveals a striking contrast between both datasets in terms of the *overall* proportions of double object (DO) versus prepositional dative (PD) instances. On the other hand, it will be shown that the alternation biases of individual dative verbs are actually quite consistent across both the corpus and the experimental data, provided these are measured in a way which evaluates the observed frequencies for individual verbs against the overall observed frequencies in the respective datasets.

1. Introduction

Linguistic studies of argument structure alternations often rely on observations about lexical *verb bias* or *verb disposition* as evidence for claims about

1. The first author is affiliated with the Linguistics department and the research unit Contragram at Ghent University, the second author with the Experimental Psychology department at Ghent University. The research of the first author was funded by the Ghent University Research Fund (contract no. GOA B/05971/01, 'Meaning in-between structure and the lexicon'). We would like to thank the volume editors as well as two anonymous reviewers for their helpful suggestions. The usual disclaimers apply.

schematic semantic contrasts between the "competing" constructions under investigation. For instance, with regard to the English dative alternation, it has been observed that verbs such as *refuse* and *deny* prefer the double object (DO) construction over the so-called prepositional dative (PD) construction (i.e., the *to*-dative), and this observation is often mentioned in support of the semantic hypothesis that the PD construction basically encodes 'caused motion' rather than 'caused reception' (e.g. Goldberg 1992, 1995, see section 2 below for further elaboration).[2] Sokolova, Lyashevskaya, and Janda (this volume) report on the lexical biases of *gruzit'* 'load' and its prefixed forms in the locative alternation in Russian.

Recently, the concept of verb bias has drawn a fair amount of attention in research on syntactic priming, too.[3] Priming effects have been shown to be sensitive to the verb-specific preferences of both *target verbs* (verbs which are strongly biased towards one of the alternating constructions are less responsive to priming of the other construction, Gries 2005) and *prime verbs* (the effect is stronger if the prime consists of a verb used in a construction it is biased *against*, Jaeger and Snider 2007). The subcategorizing preferences of individual verbs have been shown to be relevant to other psycholinguistic issues as well, such as the processing of sentences with temporary syntactic ambiguities or the ability to reproduce sentences correctly (see, e.g., Trueswell and Kim 1998; Lombardi and Potter 1992; Wilson and Garnsey 2008).

All of this raises the important question of how to measure verb-specific constructional preferences in a reliable way. The present article offers a detailed comparison of the frequency data about the Dutch dative alternation reported in Colleman's (2009) corpus-based investigation with

2. Other labels have been used for these constructions: in Goldberg's work and other recent construction grammatical analyses, for instance, the double object construction is often simply labeled *ditransitive* construction. In psycholinguistic studies, the prepositional dative is sometimes labeled *prepositional object* (PO) construction. Note that our use of the term *dative* in this context (as in "dative" constructions, the "dative" alternation, etc.) follows standard linguistic practice and should not be misunderstood as referring to constituents which are overtly marked for dative *case*. Neither Dutch nor English are morphological case languages.
3. *Syntactic* or *structural priming* refers to the tendency of speakers – first demonstrated in Bock (1986) – to repeat syntactic structures across otherwise unrelated utterances, i.e. to reuse structures from the (immediately) preceding discourse.

the results from an experimental priming study by Bernolet (2008), focussing on the evidence provided by these studies about the alternation biases of individual alternating verbs. The Dutch dative alternation is illustrated in (1) below. Just like in English, a variety of verbs of giving – as well as verbs from a number of other, semantically related verb classes such as verbs of sending or verbs of telling – can be used in either of two "competing" dative constructions in Dutch, namely (a) a double object construction with two zero-marked NP objects coding the recipient and theme participants, respectively and (b) a prepositional dative construction with a zero-marked theme object and the recipient marked by the preposition *aan* (henceforth: the *aan*-dative).[4] *Aan* is of course relevantly similar to English *to* in this respect (see Colleman and De Clerck 2009 for a detailed semantic comparison of the English *to*-dative and the Dutch *aan*-dative and Colleman 2010a for a collexeme analysis of the *aan*-dative).

(1) a. *De man heeft zijn broer een boek gegeven |*
the man has his brother a book given
verkocht | beloofd | aangeboden.
sold promised offered
'The man has given/sold/promised/offered his brother a book.'

b. *De man heeft een boek aan zijn broer*
the man has a book to his brother
gegeven | verkocht | beloofd | aangeboden.
given sold promised offered
'The man has given/sold/offered/promised a book to his brother.'

As will be illustrated below, the dative alternating verbs of Dutch display markedly different constructional preferences. The paper is structured as follows. Section 2 further sets the stage for the rest of the study, providing a discussion of the way in which the observed alternation biases of (semantically coherent subclasses of) dative alternating verbs have been

4. There are a number of constructions with other prepositions in Dutch which display a certain degree of functional overlap with the double object construction as well, but the *aan*-dative is the most systematic prepositional alternant. In construction grammar terms, the *aan*-dative can be considered the most prototypical daughter construction of the Dutch PD construction. See Van Belle and Van Langendonck (1996) and Colleman and De Clerck (2009) for further elaboration.

used as semantic evidence in studies of the dative alternation. Then, Sections 3 and 4 present the design and results of the corpus-based investigation by Colleman (2009) and Bernolet's (2008) experimental priming study, respectively. Section 5 offers a detailed comparison of the results from both investigations, which will involve a *distinctive collexeme analysis* (Gries and Stefanowitsch 2004) of both datasets. Finally, Section 6 explores the implications of our findings for current and future research on the nature and effects of lexical alternation bias while Section 7 concludes.

2. Theoretical background: Alternation biases as linguistic evidence

The dative alternation has been looked at from a myriad of theoretical perspectives, but a major research strand is concerned with the elucidation of the hypothesized subtle semantic contrasts between the DO and PD constructions. That is, many authors take the alternation to be a basically *semantic* phenomenon, associating the two constructions with related but not identical meanings. Such an approach is in accordance with the general principle assumed by most cognitive and functional approaches to grammar that, typically, in human language, a difference in grammatical form signals a difference in meaning (see e.g. Goldberg's 'Principle of No Synonymy', 1995: 67). For their evidence, semantic studies of verb alternation phenomena often rely on observations about lexical alternation bias, i.e., the preferences of verbs with particular lexical semantic characteristics for one of the "competing" constructions over the other(s) are taken as clues to the semantic differences between the constructions under investigation. For instance, an often-quoted observation about the English dative alternation – briefly mentioned in the introduction – concerns the behaviour of so-called verbs of prevention of possession such as *refuse*, *deny*, and *cost*. Goldberg (1992) reports a difference in acceptability between the DO clauses in (2a) and (3a) and their PD alternatives in (2b) and (3b), which fits in with – and as such lends added proof to – her general account of the semantic contrast between the DO and PD constructions in terms of 'caused reception' versus 'caused motion'.

(2) a. She refused Joe a raise.
    b. *She refused a raise to Joe. (Goldberg 1992: 62)

(3) a. His mother denied Billy a cake.
    b. *His mother denied a cake to Billy. (Goldberg 1992: 62)

While Joe in (2) and Billy in (3) can be construed as the projected recipients of a raise and a cake, respectively, they can hardly be construed as the goal at the end of a spatio-temporal path traversed by a raise or a cake: "[E]xpressions involving verbs of refusal (e.g., *refuse*, *deny*) cannot occur with prepositional paraphrases because they are not readily understood in terms of caused motion" (Goldberg 1992: 69). Similar statements about this class of verbs are to be found in Gropen et al. (1989), Panther (1997), and Krifka (2004), *inter alia*. All of these authors relate the observed incompatibility of *refuse* etc. with the PD construction to the purported 'caused motion' semantics of the latter. Colleman (2009: 594–596) presents a number of additional examples of semantic hypotheses about the English and/or Dutch dative alternation which build on observations about the constructional preferences of selected dative verbs.

When checked against large corpora of natural language use, such introspection-based observations often turn out to be overstated. Corpus-based work by Manning (2003) and Stefanowitsch (2006) has shown convincingly that many of the verb-structure combinations which are claimed to be ungrammatical in the linguistic literature are not actually impossible but just represent varying degrees of "rareness", so that all kinds of observations about constructional preferences should be rephrased as statistical generalizations rather than absolute constraints. In the case of *refuse* and *deny*, counterexamples to the above introspection-based observations can be found quite easily: Colleman and De Clerck (2009: 24) report figures of 107 DO versus 21 PD examples for *refuse* and 546 DO versus 118 PD examples for *deny*, out of samples of 3,000 randomly selected occurrences of each of the two verbs from the British National Corpus. The attested examples in (4) illustrate the very verb-structure combinations which are labeled ungrammatical in (2) and (3) above.

(4) a. "You're just being difficult," she said crossly. "I can not *refuse* entry to the owners. They were given the right to visit, but they don't have to produce passports. Nor do their wives or husbands." [BNC-BP91732]

b. Cardinal Cullen succeeded in getting three of the four religious provinces of Ireland to *deny* the sacraments to the Fenians. [BNC-A07607]

On the basis of such counterexamples to the introspection-based observations about verb-structure incompatibility posited in earlier studies, Bresnan et al. (2007) conclude that the semantic hypotheses about the dative alternation which these observations are generally taken to support *cannot* be

upheld (also see Bresnan and Nikitina 2009). This conclusion, however, amounts to throwing out the baby with the bathwater. Bresnan and Nikitina are of course right in so far as examples such as (4a) and (4b) show that the English dative alternation cannot be a simple matter of, for instance, 'caused reception' versus 'caused motion', but must involve *other* factors as well. Still, even if *refuse* and *deny* are thus shown to be compatible with both constructions, the skewed distributions of DO and PD instances attested with both verbs in the BNC data represent relevant facts of language in themselves.

One of the basic shared tenets of constructionist, usage-based approaches to language is the idea that grammatical items such as the DO and PD constructions are associated with semantic representations of their own which arise through generalization on the basis of encountered instances (see Goldberg 2006: Chapters 4 and 5, *inter alia*). From this usage-based perspective, it is natural to assume that, if two such abstract constructions differ in meaning, this should be evident from their occurrence in natural language in combination with different kinds of verbs filling their V-slots, i.e. not only from the *ranges* of verbs which can fill their V-slots but also from the (relative) *frequency* with which this occurs. If a particular verb is eligible for use in say both the DO construction and the PD construction, but in actual language use turns out to occur far more frequently in the former construction than in the latter, or vice versa, this is a relevant observation which may serve as lexical evidence in discussions of the semantic relation between the constructions in question. Needless to say, such observations become increasingly significant as more verbs with relatively similar lexical semantics are found to display the same alternation biases. Hence, the more verbs are included in the investigation, the more accurate the ensuing observations will be, as it becomes less and less likely that counterexamples to the advanced semantic generalizations are overlooked (a danger which looms large in introspection-based studies, as these are usually based on small samples of verbs only). This is the path chosen in the corpus-based study that will be presented in the next section.

3. Corpus-based data on alternation bias

Colleman's (2009) corpus-based investigation of the Dutch dative alternation starts out from an exhaustive set of all verbs quoted in grammars and earlier studies of the indirect object constructions of Dutch as being able to occur in the DO construction illustrated in (1a) above and/or the PD

construction with *aan* illustrated in (1b) above. All occurrences of these verbs were automatically selected from a 9 million word sample of the newspaper component of the CONDIV corpus of Present-day Dutch (Grondelaers et al. 2000) and were then manually filtered and analyzed. While the database resulting from this first sub-phase of the investigation already contained large numbers of relevant instances, the data gathering phase was complemented with a second sub-phase, in which two corpus samples of 0.5 million words each were manually skimmed to list all DO and PD instances with verbs which were not included in the original set – i.e., which had not been signalled as potential indirect object verbs in the literature before – so as to reduce the risk that potentially relevant cases were missed. Finally, all instances of these additional verbs were automatically extracted from the remaining 8 million words of the corpus sample, and then manually analyzed, in order to arrive at their total DO and PD frequencies as well. The database resulting from these two sub-phases of the investigation includes 11,116 instances of the DO and 4,949 instances of the PD construction, featuring 252 dative verbs. This set includes highly frequent verbs with several hundreds or even thousands of relevant instances in the investigated corpus sample (e.g. *geven* 'give': 2,461 DO and 939 PD instances; *bieden* 'offer': 371 DO and 179 PD instances; *vertellen* 'tell': 404 DO and 123 PD instances) as well as verbs which occur very infrequently overall and/or which are very infrequently used in the relevant constructions in the investigated corpus sample (e.g. *verkondigen* 'proclaim': 1 DO instance and 2 PD instances; *voorbehouden* 'reserve': one instance of each; *toegooien* 'throw towards': 1 DO instance and no PD instances).[5]

The overall distribution of 11,116 DO instances (=69.2% of the relevant total) versus 4,949 PD instances (=30.8%) in this database is quite similar to the figures typically reported in corpus-based studies of the dative alternation in *English*. The multifactorial corpus investigation by

5. The frequency counts in Colleman (2009) exclude passive instances (i.e., examples such as *De boeken werden* (*aan*) *hem gegeven* 'The books were given (to) him') but include instances with clausal direct objects (e.g. *Hij vertelde hem dat de winkel gesloten was* 'He told him that the store was closed'). We refer to Colleman (2009: 600–601) for a brief motivation of these and other coding decisions. Obviously, an important prerequisite for any quantitative investigation of this kind is a *formal* definition of the construction(s) to be investigated, i.e. a set of formal criteria which enable the researcher to decide in a consistent way which of the real language examples encountered in the corpus are actual occurrences of the constructions under investigation (and, crucially, which are not).

Bresnan et al. (2007), for instance, is based on a dataset of 2,360 dative observations from the three-million-word Switchboard corpus of recorded telephone conversations, viz. 1,859 (=78.8%) DO instances and 501 (=21.2%) PD instances. For the one-million-word ICE-GB, Ozón (2009) reports 587 (=68.7%) DO instances vs. 267 (=31.3%) PD instances (excluding instances of non-alternating verbs). Clearly, in Dutch as well as in English, the overall distribution of DO and PD tokens in large corpora of naturally occurring language is skewed towards the DO construction.[6]

Needless to say, the 252 verbs in Colleman's (2009) database display markedly different alternation biases, ranging from an over 95% preference for the DO (e.g. *opleveren* 'yield, earn': 284 DO vs. 3 PD instances; *leren* 'teach': 215 DO instances vs. a single PD instance) to an equally large reverse preference (e.g. *afstaan* 'hand over, cede': 65 PD instances, no DO instances in the corpus sample; *uitdelen* 'distribute': 39 PD instances vs. a single DO instance). On the basis of the observed verb-specific distributions of DO and PD instances, Colleman (2009: 602–609) presents a number of empirically valid generalizations about the kinds of verbs preferring the DO construction over the PD construction and vice versa (starting out from the results of a distinctive collexeme analysis of the observed frequency data rather than from the raw DO and PD proportions, see section 5 below for further details). For instance, with regard to the dative subclass of communication verbs, it is observed that "addressee-oriented" verbs such as *leren* 'teach', *aanraden* 'advise', *verzekeren* 'assure', *wijsmaken* 'make believe', etc. – which denote situations in which the sender of the message clearly wants to influence the receiver's future actions – display a stronger preference for the DO construction than verbs such as *meedelen* 'communicate', *laten weten* 'let know', *signaleren* 'signal, draw attention to', etc., which denote a more neutral transfer of a message. This observation is in line with existing hypotheses about the stronger

6. The even larger proportion of DO instances in the English database compiled by Bresnan and colleagues as compared to the Dutch data in Colleman (2009) might be partly due to the different *modes* of language represented by the corpora used, viz. the Switchboard corpus of recorded telephone conversations versus the newspaper component of the CONDIV corpus. However, there are at present no detailed frequency data available on indirect object constructions in spoken varieties of Dutch, so that this claim cannot be tested. It should also be noted that the PD frequencies reported by Ozón (2009) include *for*-datives as well as *to*-datives. Since the former prepositional construction is quite infrequent, this does probably not make much of a difference.

profiling of the involvement of the recipient/beneficiary/addressee participant in the DO construction (see, e.g., Wierzbicka 1986, Langacker 1991: 357–360, or, specifically on Dutch, Van Belle & Van Langendonck 1996). Another generalization, which will be discussed in more detail in section 5 below, concerns the consistently over-average PD-preference of particle verbs with *af* 'off', *door* 'through', *over* 'over' or *uit* 'out'. First, however, the next section presents the results from an experimental study of the Dutch dative alternation reported in Bernolet (2008).

4. Experimental data on alternation bias

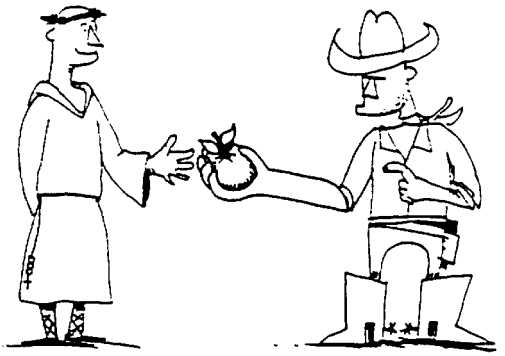
Bernolet (2008: Chapter 5) aims to investigate the effects of lexical alternation bias on the strength of structural priming through a series of picture description experiments involving Dutch dative verbs (also see Bernolet and Hartsuiker 2010 for a condensed discussion of one of these experiments). To this end, 18 test verbs with varying alternation biases were selected from a preliminary version of Colleman's (2009) database, cf. the list in (5).

(5) *aanbieden* 'offer', *betalen* 'pay', *bezorgen* 'deliver, cause', *doorgeven* 'pass, hand (on)', *geven* 'give', *laten zien* 'show' (lit. *let see*), *meegeven* 'give with, send with', *overhandigen* 'hand', *schenken* 'give (as a present)', *schrijven* 'write', *teruggeven* 'give back', *tonen* 'show', *uitreiken* 'give out, issue', *verklappen* 'let out, spill', *verkopen* 'sell', *voorleggen* 'present, submit', *voorlezen* 'read out', *voorstellen* 'suggest, introduce'

The selection includes highly frequent (e.g. *geven* 'give', *bezorgen* 'cause, deliver', *verkopen* 'sell') as well as relatively infrequent verbs (e.g. *teruggeven* 'give back', *verklappen* 'let out, spill'). As for their alternation bias, the selected verbs range from an over 90% preference for the DO construction to a similar preference for the PD construction in Colleman's corpus data. Further details are provided in subsection 4.3 below.

4.1. Experiment 1

In a first experiment, which served as a pretest for the actual priming study, 943 undergraduate students at Ghent University, all native speakers of Dutch who participated on a voluntary basis, were asked to provide a written description of a target picture presented to them on a slip of paper.



DOORGEVEN

Figure 1. Example of a target picture for *doorgeven* 'pass, hand (on)'. Intended result: *De cowboy geeft de monnik de/een appel door* 'The cowboy passes the monk the/an apple' or *De cowboy geeft de/een appel door aan de monnik* 'The cowboy passes the/an apple to the monk'

For each of the 18 selected verbs, 3 different target pictures were constructed. The pictures all showed line drawings of dative actions involving persons and objects with the roles of agent, theme, and recipient. In all pictures, the recipient participant was depicted on the left, the agent on the right, and the theme in-between the agent and recipient participants. For each of the 54 target pictures, a different combination of persons and objects was used. Beneath the actions on the pictures, a dative verb was printed, see Figure 1 for an example of a target picture for *doorgeven* 'pass, hand (on)'. Above the actions, a brief instruction was printed ('write a short sentence that describes what you see on the picture, using the verb that is printed beneath the picture'). The participants were tested in a classroom setting. Each participant received only one target picture, and the sheets of paper were collected as soon as the participants had written down a sentence. For each target verb, at least 50 descriptions were collected.

Of the 943 elicited sentences that were collected, 131 represented the DO construction (13.9%) and 487 represented the PD construction with *aan* (52.1%). 325 responses (34.8%) were coded as 'Others': since the participants were not instructed to produce either the DO or the PD construction, they regularly used monotransitive constructions, constructions with other prepositions than *aan*, passives, etc. in their answers.

4.2. Experiment 2

The second experiment reported in Bernolet (2008: Chapter 5) is an actual syntactic priming experiment. Thirty undergraduate students at Ghent University (28 females and 2 males, all native speakers of Dutch) were paid to take part and a female undergraduate student acted as confederate. The set with critical stimuli for the participants contained 54 pictures showing line drawings of dative actions with one of the 18 test verbs printed beneath (i.e., the same pictures that were used in Experiment 1). For each of these pictures three prime sentences were constructed (one for each prime condition): a prime sentence using the DO construction, a prime using the PD construction with *aan* and a baseline prime sentence using an intransitive or monotransitive construction. In the DO and PD conditions, the prime verb was always identical to the verb that was depicted on the corresponding target picture; in the baseline condition a transitive or intransitive verb was used in the prime sentence. Apart from the critical pictures, 108 non-critical pictures were selected as fillers. The fillers either showed pictures of intransitive (e.g., a weeping cowboy) or transitive actions (e.g., a nun chasing a swimmer). Prime sentences were constructed for the filler pictures as well. In 63 filler pairs the same verb was used in prime and target, in the remaining 45 filler pairs a different verb was used. Consequently, the same verb had to be used in prime and target constructions in half of the trials (critical + filler trials). Additionally, 162 pictures were selected for the verification set of participant and confederate. These pictures were used for a verification task that was used to mask the real purpose of the experiment.

The participants were tested in groups of two in a dialogue experiment (Bernolet, Hartsuiker & Pickering 2007; Branigan et al. 2000). Both dialogue partners, one of whom was the confederate, took turns in describing pictures that appeared on the screen of their computers. They were instructed to listen and react to their dialogue partner's descriptions by pressing '1' if the description matched the picture that was simultaneously presented on their computer screen or '2' if the description and the picture did not match. Instead of describing pictures, the confederate read prime sentences from the screen of her computer. These prime sentences were presented in three counterbalanced lists. In each of these lists the primes were presented equally often in the three priming conditions (DO-prime, PD-prime, baseline) and across all participants every target picture was presented equally often in each of the three conditions. Each verb was used three times in each list, once in each priming condition.

Across all priming conditions, the participants in Experiment 2 produced 432 DO responses (26.7%), 1,058 PD responses (65.3%) and 130 other responses (8%). In the baseline condition, the distribution of responses was 106 DO instances (19.6%), 364 PD instances (67.4%) and 70 others (13%).

4.3. Overview

The second column of Table 1 below presents the exact DO and PD corpus frequencies in Colleman's (2009) database of the 18 verbs that were selected for Bernolet's (2008) study. As can be seen in the third column, the alternation biases, in terms of the raw proportions of DO occurrences in the total number of relevant instances for each verb, range from 10.5% (*uitreiken* 'give out, issue') to 93.1% (*bezorgen* 'deliver, cause'), with the remaining verbs covering the full spectrum.

Table 1. Observed DO and PD frequencies and alternation biases for the 18 test verbs in Colleman's (2009) corpus data and Bernolet's (2008) experimental data (verbs listed in ascending order of DO-preference in the corpus data)

	corpus data		experimental data	
	observed frequencies DO:PD	alternation bias (% DO-datives)	observed frequencies DO:PD	alternation bias (% DO-datives)
uitreiken 'give out, issue'	2:17	10.5%	3:64	4.5%
doorgeven 'pass, hand (on)'	9:61	12.9%	1:70	1.4%
verkopen 'sell'	39:204	16.0%	4:63	5.8%
voorleggen 'present, submit'	36:123	22.6%	13:53	19.7%
schrijven 'write'	36:92	28.1%	1:29	3.3%
overhandigen 'hand'	34:70	32.7%	16:55	22.5%
betalen 'pay'	63:111	36.2%	6:6	50.0%
schenken 'give (as a present)'	71:100	41.5%	17:60	22.1%
teruggeven 'give back'	24:26	48.0%	9:52	14.8%
voorstellen 'suggest, introduce'	72:55	56.7%	1:63	1.6%
verklappen 'let out, spill'	4:3	57.1%	5:37	11.9%
tonen 'show'	70:31	69.3%	13:53	19.7%
voorlezen 'read out'	7:3	70.0%	9:36	20.0%
geven 'give'	2461:939	72.4%	11:66	14.3%
laten zien 'show'	59:20	74.7%	26:41	38.8%
aanbieden 'offer'	205:68	75.1%	37:32	53.6%
meegeven 'give with, send with'	71:15	82.6%	22:42	34.4%
bezorgen 'deliver, cause'	335:25	93.1%	43:29	59.7%
TOTAL	3598:1963	64.7%	237:851	21.8%

The two rightmost columns provide the same information for the experimental study. We collapsed the results from Experiments 1 and 2, since these basically involve the same picture description task (albeit once in written and once in spoken mode). However, from the second experiment, we only included the responses in the *baseline* condition: in this way, the selected data only include target descriptions which were produced *without* a DO or PD prime in the immediately preceding context, so that the results are not contaminated by immediate priming effects.[7] A two-tailed t-test for paired observations corroborated that the alternation biases of the 18 verbs in Experiment 1 on the one hand and in the baseline condition of Experiment 2 on the other do not differ significantly, so that it is justified to collapse these data ($t(17) = 0.94$, $p > .1$).[8] In all, the number of relevant dative responses in the two picture description tests totals 1,088, of which 237 (=21.8%) are DO-datives and 851 (=78.2%) are PDs, as shown in the bottom row of Table 1. For the sake of convenience, we will sometimes refer to the data from the picture description experiment in what follows, when what is actually meant are the data from the two parts of the investigation collapsed in the right-hand columns of Table 1.

It should be added that the reported PD frequencies concern the *aan*-dative only: in the corpus data as well as in the picture description responses, some of these verbs were sporadically attested in three-participant constructions with other prepositions as well, but such instances were discarded (an instance is *schrijven* 'write', which occurs with *naar* 'to' as well as *aan*, see Colleman and De Clerck 2009 on the relation between these prepositions).

7. This may seem an unnecessary precaution given that there are equal numbers of DO prime and PD prime test conditions, which might be thought to neutralize the possibly confounding effects of syntactic priming on the overall distribution of DO and PD datives. However, as will be briefly discussed in section 6, the DO construction was found to display a *stronger* priming effect than the PD construction, so that it is best not to include these data in our main analyses.
8. For this test, the alternation biases were simply measured as the proportion of DO-datives in the total number of dative responses for each verb. If, however, the alternation biases are computed as in Bernolet & Hartsuiker (2010), viz. as the log of the number of DO responses + 1 divided by the number of PD responses + 1, the result is relevantly similar ($t(17) = 1.58$, $p > .1$). See Bernolet and Hartsuiker (2010: 457, note 2) for a brief motivation of this measure of alternation bias.

5. A comparison of the two datasets

5.1. Introduction

Even a quick glance at Table 1 suffices to reveal a striking difference between the two datasets in terms of the overall distribution of DO and PD instances. Note that for this particular selection of 18 verbs the overall DO preference revealed by the corpus data is slightly lower compared to the overall frequencies for *all* 252 verbs in the database reported in section 3 above, viz. 64.7% rather than 69.2% DO instances. But this is obviously still a very clear overall DO preference, so that the corpus results are clearly at odds with the experimental results in which the DO responses constitute only slightly over one-fifth of the total number of relevant dative responses (21.8%).

Turning to the attested DO and PD frequencies of the individual test verbs, it can be observed that, for 17 out of the 18 investigated verbs, the proportion of DO instances is lower in the experimental data than in the corpus data. The single exception is *betalen* 'pay' (36.2% DO-datives in the corpus material, 50% DO-datives in the experimental data), but note that the overall number of relevant responses for this verb in the picture description experiment is very small, precluding firm conclusions about its alternation bias in the experiment.[9] If the raw proportion of DO-datives is taken as the measure of alternation bias, several verbs switch sides, so to speak, and appear distinctly DO-biased from the corpus data but distinctly PD-biased from the picture description data: instances include *geven* 'give' (72.4% and 14.3% DO-datives, respectively) and *tonen* 'show' (69.3% and 19.7% DO-datives, respectively). However, it can also be observed that, in general, the verbs with the highest proportions of DO-datives in the corpus data also display the highest proportions of

9. *Betalen* 'pay' is in fact the only verb for which the experiment produced a majority of 'Other' responses. This is probably due to the fact that the depicted theme in all three of the *betalen* target pictures was a couple of bank-notes in the giver's hand. It seems that these did not "stand out" enough to be perceived as a full-fledged third participant in the event to be described, since the majority of responses were simple monotransitives of the kind *De matroos betaalt de schilder* 'The sailor pays the painter'. It should be added that in corpus data as well, *betalen* tends to occur far more frequently in two-participant than three-participant constructions (e.g. Colleman 2006: 643 reports 1,294 mono-transitive instances in a set of 1,500 randomly collected *betalen* occurrences from the CONDIV corpus).

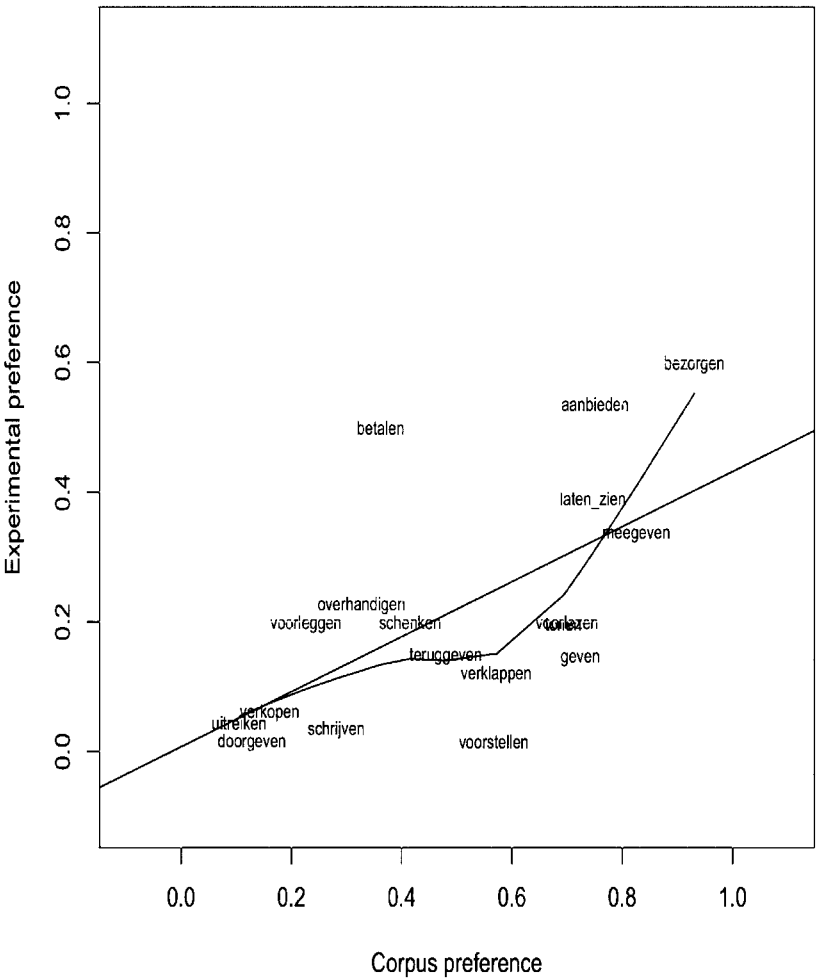


Figure 2. DO proportions of the 18 test verbs in the corpus data and the experimental data

DO-datives in the experimental data. In fact, there is a significant positive linear correlation between the alternation biases from both sources measured in terms of the raw DO proportions ($r = .606$, $N = 18$, $p < .05$), indicating that both variables tend to increase or decrease together. This correlation is shown graphically in Figure 2.

In the remainder of this section, we first explore a number of possible reasons for the observed contrast in overall DO and PD proportions (subsection 5.2) and then turn to a closer analysis of the results for the individual test verbs (subsections 5.3 and 5.4).

5.2. The overall proportions

Before we move on, it should be observed that the strong overall PD preference attested in Bernolet (2008) was not entirely unanticipated. Other experimental priming studies of the Dutch dative alternation have reported an overall bias towards the prepositional dative as well: 54% PD responses across all priming conditions in Hartsuiker et al. (2008), 73% PD responses across all priming conditions in Schoonbaert, Hartsuiker, and Pickering (2007), 71% PD responses in the baseline condition in Melinger and Dobel (2005). As for corpus data, there are no other corpus investigations of the Dutch dative alternation similar in scope to Colleman (2009), but the results from a number of earlier small-scale corpus studies point towards a clear overall preference for DO-datives as well: Kirsner (1988) counted all occurrences of the competing dative constructions in 4 contemporary Dutch novels and found 232 DO-datives (=80.8%) versus 55 PDs (=19.2%), Schermer-Vermeer (1991: 295) found 237 DO-datives (=86.8%) and 36 PDs (=13.2%) in a very similar corpus of contemporary fiction. In sum, there is a consistent predominance of the DO construction in Dutch corpus data and a consistent predominance of prepositional datives in experimental data.

Interestingly, this contrast is partly mirrored in English. As was observed in Section 3, corpus studies of the English alternation such as Bresnan et al. (2007) and Ozón (2009) report strong overall DO preferences (78.8% and 68.7% DO-datives, respectively). Experimental studies, by contrast, typically report more balanced proportions of DO and PD responses in the elicited data, with often even a slight overall PD bias. In the study by Pickering and Branigan (1998), for instance, which investigates the effects of structural priming in a series of controlled sentence completion experiments, the DO-dative was *not* the most frequently produced construction. The participants in the first experiment reported in the paper were asked to complete target sentences consisting of a subject NP followed by a verb that could be completed with DO or PD syntax (e.g. *The bus driver gave...*, *The little girl handed...*). In all 1,042 trials, i.e. across all test conditions, they produced 390 (=37%) PD completions and 303 (=29%) DO completions (next to 34% of other completions). Relevantly similar results were obtained by Corley

and Scheepers (2002) (33.3% PD and 24.8% DO), Branigan, Pickering, and Cleland (2000) (55.3% PD and 44.8% DO) and Branigan et al. (2000) (49% PD and 51% DO). So the PD scores fare better in elicited production data than in corpus data in both English and Dutch, although the contrast is even more outspoken in the latter language than in the former.[10]

Roland and Jurafsky (2002) address the different results yielded by different methods of calculating verb subcategorization probabilities, i.e. between subcategorization frequencies computed from corpora and those computed from psychological experiments, or between subcategorization frequencies computed from different kinds of natural-language corpora. They identify two broad sources of variation, viz. *context-based variation* and *word-sense variation*, the former of which is especially relevant to the different results obtained from corpora and production experiments. The authors stress the inherently different nature of single sentence production and connected discourse, or of "test-tube" sentences versus "wild" sentences. They observe that the proportion of *passive* clauses, for instance, is consistently higher in natural language corpora than in data from production experiments, which can be related to the agent de-emphasizing and patient topicalizing functions of the passive, since these are of course more relevant for multi-sentence discourse than for isolated sentences.

The different behaviour of the DO and PD constructions "in the wild" as compared to in controlled experimental settings can be related to discourse context effects as well. Colleman's (2009) corpus data comprise a wide variety of sentence types: DOs and PDs with pronominal as well as lexical subject and object NPs, in main clauses as well as in (finite or infinitival) subordinate clauses, in clauses with canonical word order as well as in clauses with a fronted object or with other alternative word orders, and so on. By contrast, the isolated sentences produced in the picture description experiment are all of the same type, i.e., simple main clauses with lexical NP subjects and objects (e.g. *de cowboy* 'the cowboy', *de kok* 'the cook', *een appel* 'an apple') in the canonical word order (i.e., subject-verb-indirect object-direct object for the DO construction and subject-

10. In this regard, it is also interesting to note that in the first picture description experiment reported in Schoonbaert, Hartsuiker, and Pickering (2007), which tested priming effects in the *L2 English* of native speakers of Dutch, the percentage of PD target descriptions was closer to the percentages observed for the PD in Dutch than to the percentages typically observed for the PD in L1 English, viz. 74.7% PD descriptions vs. 25.3% DO descriptions.

verb-direct object-indirect object for the PD).[11] The dative alternation is well-known to be partly driven by discourse factors such as the pronominality and definiteness of the object NPs and the discourse-givenness and specificity of the object referents, which are all parameters contributing to the relative "topicality" of the theme and recipient participants. In case of a (highly) topical recipient participant (and a theme participant with low topicality), the DO construction is the preferred option, since this construction presents the recipient in the immediately postverbal position, *before* the theme participant; in the reverse case the PD construction with its unmarked theme-recipient order will be the preferred option. In this way, the constructional choices of speakers are affected by principles of discourse cohesion (see Collins 1995, Goldberg 1995, Gries 2003, Bresnan et al. 2007, *inter alia*).

Table 2 presents the recomputed corpus frequencies in Colleman's (2009) database when just a single of these parameters is controlled for, viz. the *pronominality* of the recipient. The figures in Table 2 only include observed DO and PD instances in which the indirect object is realized as a lexical rather than pronominal NP, just like in the isolated sentences from the picture description experiments, thus factoring out the impact of one of the parameters known to boost the use of the DO construction in connected discourse.

As can be seen from a comparison of the bottom rows in Tables 1 and 2, this simple exclusion of all instances with pronominal indirect objects results in a 7.6% drop in the overall proportion of DO datives. A two-tailed *t*-test for paired observations indicates that the alternation biases on the left-hand side of Table 1 differ significantly from the alternation biases in Table 2 ($t(17) = 3.35$, $p < .01$).[12] In this way, the observed contrast between the overall distributions of the DO and PD constructions in

11. It should be added that the participants in the two parts of the experiment also produced a relatively small number of "shifted" *aan*-datives, i.e. with the non-canonical order of the prepositional indirect object *before* the direct object (e.g. *De soldaat geeft aan de monnik een boek* 'The soldier gives to the monk a book'). In the written picture description test, 33 of the 943 elicited clauses were of this type (=3.5%). These shifted *aan*-datives were classified as 'Others' in Bernolet (2008). If they would have been counted as genuine instances of the *aan*-dative, the overall preference for this construction would have been even slightly higher. See Hartsuiker and Kolk (1998) for a priming experiment in which the shifted PD construction is treated as a full-fledged third option, on a par with the DO construction and the canonical PD.
12. Again, the result of the test is the same if the alternation biases are computed as the log of the number of DO responses + 1 divided by the number of PD responses + 1 rather than as the raw proportion of DO responses ($t(17) = 5.37$, $p < .001$).

Table 2. Observed DO- and PD-frequencies and alternation biases for the 18 test verbs in Colleman's (2009) corpus data, excluding cases with pronominal indirect objects

	Corpus frequencies excluding [+pronominal] indirect objects	
	observed frequencies DO:PD	alternation bias (% DO-datives)
uitreiken 'give out, issue'	0:16	0%
doorgeven 'pass, hand (on)'	4:58	6.5%
verkopen 'sell'	23:197	10.5%
voorleggen 'present, submit'	20:114	14.9%
schrijven 'write'	17:90	15.9%
overhandigen 'hand'	19:70	21.4%
betalen 'pay'	47:109	30.1%
schenken 'give (as a present)'	37:97	27.6%
teruggeven 'give back'	17:25	40.5%
voorstellen 'introduce, suggest'	56:48	53.9%
verklappen 'let out, spill'	0:3	0%
tonen 'show'	35:30	53.9%
voorlezen 'read out'	0:3	0%
geven 'give'	1680:829	66.9%
laten zien 'show'	31:19	62%
aanbieden 'offer'	134:68	66.3%
meegeven 'give with, send with'	50:15	76.9%
bezorgen 'deliver, cause'	247:23	91.5%
TOTAL	2417:1814	57.1%

corpus versus experimental data is at least partly due to context-based variation.

Note that there may be further context-based parameters at work besides the effects of discourse cohesion mechanisms in natural discourse. Unlike the DO construction, the PD features an overt marker of recipient function, in the form of the preposition *aan* (or, in English, *to*). Kirsner (1988: 290) raises the interesting idea that, under some circumstances, speakers may opt for the PD construction "from a desire to achieve maximum clarity and explicitness". It could very well be the case that in the laboratory context of a picture description experiment speakers feel more compelled to

produce descriptions of maximum perceived accuracy and explicitness than they do in natural settings. We leave it to further research to test the possible effects of such parameters.[13]

5.3. DO-biased and PD-biased verbs

In the preliminary discussion of the figures in Table 1 in subsection 5.1 above, we noted that the verbs with the highest proportions of DO-datives in the corpus data also tend to display the relatively highest DO proportions in the experimental data. This correspondence between the alternation biases displayed by the investigated verbs in the two datasets can be further illustrated through an application of *distinctive collexeme analysis*, one of the family of the so-called collostructional methods developed in a series of papers by Gries and Stefanowitsch (Stefanowitsch and Gries 2003, 2005; Gries and Stefanowitsch 2004), which are aimed at determining the degree of association between abstract constructions and the lexical items filling their constructional slots, or between lexical items occurring in various slots of the same construction. For a full explanation and justification of this methodology we would like to refer to Gries and Stefanowitsch (2004), but, very briefly, distinctive collexeme analysis tests the degree of association between two or more "competing" constructions C_1, C_2, etc. and the various lexemes occurring in a particular slot of these constructions, on the basis of the co-occurrence frequencies of lexeme x and C_1, x and C_2, etc. *and* the overall frequencies of C_1, C_2, etc. in the corpus. A lexeme is revealed by this test to be significantly attracted to one of the constructions under investigation if its observed frequency in that construction significantly exceeds the frequency expected on the basis of the overall distributions. The distributional statistic used for this is the Fisher Exact test, a decision which is motivated in statistical terms in Stefanowitsch and Gries (2003: 217–219). If this procedure is repeated

13. Another kind of "contextual" variation, not mentioned by Roland and Jurafsky (2002), is between-speakers variation, for instance along regional lines. Note that, while the corpus data are derived from a newspaper corpus with equal proportions of Belgian and Netherlandic text material, the participants in the picture description experiment were all students at Ghent University, i.e. native speakers of Belgian Dutch. The possible effects of regional (or other sociolinguistic) parameters on the dative alternation in Dutch have never been systematically investigated, although the DO construction is known to have a slightly different semantic range in Belgian vs. Netherlandic varieties of the language (see, e.g., Colleman 2010b).

for all lexemes occurring in the investigated slot of either of the constructions in the corpus, the outcome is a list of so-called *distinctive collexemes* for each of the examined constructions, i.e. the lexemes with a significantly above-average preference for that construction over the other construction(s). Moreover, these distinctive collexemes can be ranked by using the outcome of the distributional test as a measure of distinctiveness: the *smaller* the p-value resulting from the test, the *stronger* the association between the lexeme and construction in question. For completeness' sake, it should be added that the test does not take into account the overall frequency of the investigated verbs *outside* of the investigated constructions.

The distributional generalizations in Colleman (2009) are not based on the raw DO and PD corpus frequencies of the investigated verbs, but on the results of a distinctive collexeme analysis of these frequencies. Of the 252 verbs with one or more DO and/or PD instances in the investigated sample from the CONDIV corpus, 73 verbs are revealed by this test to display a significantly ($p < .05$) above-average preference for the DO construction, and 58 verbs are revealed to display a significantly ($p < .05$) above-average preference for the PD construction. The remaining 121 verbs are neutral, i.e. their observed DO and PD frequencies do not differ significantly from the frequencies expected on the basis of the overall distribution of both constructions in the corpus sample (i.e., 11,116 instances of the DO construction and 4,949 instances of the PD construction, see section 3 above). The lefthand side of Table 3 summarizes the results of this test for the 18 verbs included in the picture description experiments: 4 of them belong to the significant collexemes of the DO-dative in the corpus material, 10 to the significant collexemes of the PD, and 4 are neutral.[14] *Bezorgen* 'deliver, cause', the top verb in the leftmost column of Table 3, occupies the fourth position in the full ranking of DO collexemes presented in Colleman (2009), after *opleveren* 'yield, earn', *wijten* 'accuse, blame', and *leren* 'teach', none of which was included in Bernolet's (2008) sample of 18 test verbs. Similarly, *verkopen* 'sell', the top verb among the PD collexemes listed in the lefthand side of Table 3, is preceded by *overlaten* 'leave, pass on' in the full list of PD collexemes. See Colleman (2009: 602) for an overview of the thirty most strongly distinctive collexemes of both constructions.

14. It should be added that Table 3 follows a convention which is advocated by Gries and Stefanowitsch in later work, viz. the use of *log-transformed p*-values as a measure of collostruction strength (see e.g. Gries, Hampe, and Schönefeld 2005: 648). These can be interpreted as follows: distinctiveness $> 3 \Rightarrow p < .001$; distinctiveness $> 2 \Rightarrow p < .01$; distinctiveness $> 1.30103 \Rightarrow p < .05$. The larger the distinctiveness value, the stronger the attraction.

Table 3. Distinctive collexemes of the DO and PD constructions in the corpus data and the experimental data

Corpus data				Experimental data			
DO		PD		DO		PD	
Collexemes	Distinct.	Collexemes	Distinct.	Collexemes	Distinct.	Collexemes	Distinct.
bezorgen	28.41	*verkopen*	66.26	*bezorgen*	12.35	*doorgeven*	6.51
geven	5.56	*voorleggen*	33.13	*aanbieden*	8.72	*voorstellen*	5.76
meegeven	2.44	*doorgeven*	21.89	*laten zien*	3.09	*uitreiken*	4.23
aanbieden	1.74	*schrijven*	20.97	*meegeven*	1.93	*verkopen*	3.52
		betalen	18.66	*betalen*	1.55	*schrijven*	2.27
		overhandigen	13.76				
		schenken	13.28				
		uitreiken	6.77				
		teruggeven	2.85				
		voorstellen	2.72				
Neutral: *verklappen*, *voorlezen*, *tonen*, *laten zien*				Neutral: *geven*, *verklappen*, *voorlezen*, *tonen*, *voorleggen*, *overhandigen*, *schenken*, *teruggeven*			

The righthand side of Table 3 shows the results of a distinctive collexeme analysis of the data from the picture description experiment. Of course, in this case, the overall distribution is skewed towards the PD rather than towards the DO construction. Still, the results of both tests are similar in a number of ways. Eight out of the ten verbs which display a significantly above-average preference for either of the two constructions in the experimental data, display the very same alternation bias in the corpus data. There is only one verb which switches sides, viz. *betalen* 'pay', which is significantly attracted to the PD construction in the corpus material but which is among the significant DO collexemes in the data from the picture description experiment. However, nothing much should be made of this contrast, since, as was mentioned at the beginning of this section, the number of relevant *betalen* 'pay' instances produced in the experiment is very low. *Laten zien* 'show' is significantly attracted to the DO construction in the experimental data, but is classified as 'neutral' in the corpus data: in fact, it displays an above-average preference for the DO construction in *both* datasets, but in the corpus data, the difference between observed and expected frequencies fails to reach significance. Vice versa,

there are a number of verbs which are significantly attracted to one of the two constructions in the corpus data but which do not show a significantly above-average preference in the experimental data, such as *voorleggen* 'present, submit' and *overhandigen* 'hand' (note that the larger number of neutral verbs on the righthand side of the table is not surprising given the smaller number of relevant observations in the experimental dataset). At the very least, in such cases, the results from the experiment do not *contradict* the alternation biases computed on the basis of the corpus data.

While the sample of dative verbs included in the picture description experiments is quite small, the results provide additional support for at least one of the semantic generalizations put forward in Colleman (2009), which concerns the behaviour of *verkopen* 'sell' and of a particular subclass of particle verbs. In the data from the CONDIV corpus, particle verbs with one of the particles *af* 'off', *door* 'through', *over* 'over' or *uit* 'out' as their first element are consistently strongly attracted to the PD construction. Relevant examples include *overlaten* 'leave, pass on' (0:134), *overdragen* 'hand over, transmit' (1:84), *afstaan* 'cede, hand over, part with' (0:65), *overmaken* 'transfer, remit' (6:59), *doorgeven* 'pass on' (9:61), *uitdelen* 'distribute' (1:39), *uitleveren* 'extradite, hand over' (1:32), *doorspelen* 'pass on, leak' (3:25), *afgeven* 'hand in' (1:20) and *uitreiken* 'give out, issue' (2:17), all of which are among the thirty most strongly distinctive collexemes of the PD construction as revealed by a distinctive collexeme analysis of the observed corpus frequencies (the figures in brackets refer to the observed number of DO and PD instances in the 9 million word sample from the CONDIV corpus, respectively). By contrast, other classes of spatial particle verbs – e.g. with *toe* 'towards', *in* 'in', *voor* 'for', *om* 'around', etc. – do not display a consistent preference for the PD construction. The behaviour of the first subclass of particle verbs can be related to their lexical emphasis on the changing agent-theme relation, i.e. the particles in question seem to bring a sense of 'separation' to the meaning of the complex verb. A speaker who selects for instance the verb *afgeven* 'hand in' or *overgeven* 'hand over' to describe a possessional transfer event rather than the basic verb *geven* 'give', thereby focuses on the agent's parting with the theme, or put differently, on the theme's leaving the agent's domain. This aspect of these verbs' lexical semantics tallies better with the prepositional dative – which, in construction grammar terms, constructionally profiles the involvement of the agent and theme participants in the depicted transfer event – than with the double object construction, which constructionally profiles the involvement of all three participants (see Goldberg 1995: 48–49 for the notion of constructional profiling and see Colleman 2009: 605–609

for a more elaborate version of the argument). *Verkopen* 'sell', though not a particle verb, is relevantly similar to these verbs in that, of the four major participants in a commercial transaction event – the buyer, the seller, the goods, and the money –, it lexically profiles the seller and the goods (see e.g. the traditional frame semantic analyses of *buy* and *sell* in Fillmore 1977). As such, it lexically highlights the agent-theme relation as well, hence its preference for the PD construction in the corpus data: it is identified as the second strongest PD-collexeme overall in Colleman (2009).

What matters most in the current context is that in the data from the picture description experiment, *verkopen* 'sell' is identified as being strongly attracted to the PD construction *as well*. The same applies to the only two members of the subclass of particle verbs discussed above that were included in the experiment, viz. *doorgeven* 'pass, hand (on)' and *uitreiken* 'give out, issue'. This goes to show that the lexical alternation biases of such verbs are quite robust: even in a dataset with a strong overall bias towards the PD construction, *verkopen*, *doorgeven* and *uitreiken* still stand out as displaying a significantly *above-average* preference for this construction. As such, the data from the corpus investigation and from the psycholinguistic experiment provide converging evidence for the lexical preferences of these verbs for the PD construction. Similarly, they provide converging evidence for the lexical alternation biases of *bezorgen* 'deliver, cause', *aanbieden* 'offer' and *meegeven* 'give with, send with', which come out as distinctly DO-biased on both counts. The next section looks into a couple of verbs for which the results from the two investigations are less uniform, and explores the reasons why this might be so. This will trigger a discussion on the relation between lexical alternation bias and verbal polysemy.

5.4. A closer look on *voorstellen*, *geven* and *schenken*

The first verb to be discussed in somewhat more detail is *voorstellen*, which, as indicated in Table 3, is revealed by the distinctive collexeme analyses to be significantly attracted to the PD construction in both datasets. However, whereas it only occupies the tenth (and last) position in the ranking of significant PD-collexemes based on the corpus data, with observed DO and PD frequencies of 72 and 55, respectively, it is the second most strongly attracted PD-collexeme in the data from the picture description experiment, with a single DO response versus 63 PD responses. We will not present a detailed comparison of the exact rankings of all test verbs in both parts of the study, nor of their attained distinctiveness scores,

which, for one thing, are strongly dependent on the sample sizes. In the case of *voorstellen*, however, the contrast between both rankings is striking enough to merit closer attention. The ditransitive uses of *voorstellen* cover a quite broad semantic range. On the one hand, there is a family of uses which, in English, are rendered by means of verbs such as *suggest*, *propose* or *offer*: proposing a solution to someone, suggesting a place for lunch to someone, offering a new contract proposal to someone, etc. On the other hand, there is a family of uses which correspond to verbs such as *introduce* and *present* in English: introducing a friend to someone, presenting a new book to an audience, etc. The two broad senses tend to occur with very different types of direct object referents: mostly ideas in the former sense, mostly (but not exlusively) persons in the latter. As shown in the examples from CONDIV in (6) and (7) below, both senses occur with both DO syntax and PD syntax.[15]

(6) a. *President Clinton gaat het Congres volgende maand de*
Pres. C. goes the Congress next month the
grootste toename in defensie-uitgaven sinds de
largest increase in defense-spendings since the
Koude Oorlog voorstellen. [NRC]
Cold War propose
'Next month, President Clinton is going to propose the largest increase in defense spending since the Cold War to Congress.'

b. *De bond gaat een plan voorstellen aan de clubs*
the assoc. goes a plan propose to the clubs
uit eerste. {Wij willen scheidsrechters voortaan 5.000
from first
frank per eersteklassewedstrijd uitkeren.} [GvA]
'The association will propose a plan to all clubs in first division. {From now on, we want to award referees 5,000 francs for each first division game.}'

15. The labels in brackets indicate the exact source, NRC refers to the Dutch newspaper *NRC Handelsblad*, GvA and HBL to the Belgian newspapers *Gazet van Antwerpen* and *Het Belang van Limburg*, respectively. In (6b), the *voorstellen* verb appears in its split form, with the particle *voor* 'for' separated from its verbal base *stellen*. For an introduction to the grammatical behaviour and properties of separable particle verbs in Dutch, see Booij (2002).

(7) a. {*Jonge architecten die hun eerste huis, hun eerste project uitvoeren,*}
die willen we u voorstellen. [GvA]
them want we you present
{'Young architects who are realizing their first house, their first renovation or project,} those are the people we want to present to you.'

b. {*Robby kwam nog een keer langs, om te vertellen dat hij homo was.*}
Hij stelde zijn vriend aan haar voor.
he VERB his friend to her PART
'{Robby came by one more time, to tell her that he was gay.}
He introduced his boyfriend to her.' [NRC]

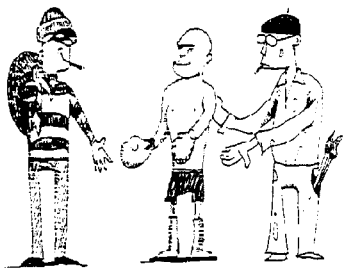
However, both groups of uses behave markedly differently with regard to the dative alternation, as shown in Table 4, which splits up the *voorstellen* instances attested in the corpus data into two broad semantic categories.[16] Clearly, the 'propose, suggest' uses prefer the DO construction over the PD construction, whereas the 'introduce, present' uses display the reverse preference (the contrast between the two observed distributions is statistically significant, Chi-square = 41.80, $df = 1$, $p < .0001$).

Now consider Figure 3, which shows the target pictures for *voorstellen* used in the experiment. In all three pictures, the depicted theme is a human participant (a boxer, a ballerina, and a clown, respectively), which naturally triggers an interpretation of the scene as an 'introduce, present' event rather than as a 'propose, suggest' event. In view of this, the stronger attraction of *voorstellen* to the PD construction in the results from the experiment as compared to the corpus data ceases to be surprising. In

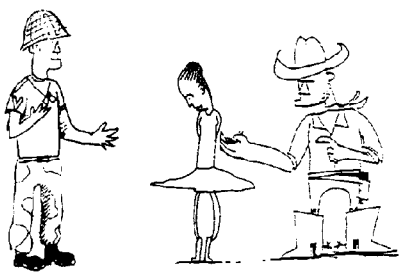
Table 4. Observed corpus frequencies for the two broad senses of *voorstellen*

	DO	PD	total
'suggest, propose'	67	21	78
'introduce, present'	3	28	31
total	70	49	119

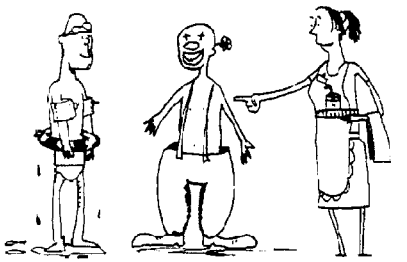
16. The frequencies in Table 4 do not add up to 127 (=the overall frequency of dative *voorstellen* mentioned in Table 1), as there is a small number of instances which resist straightforward semantic classification.



VOORSTELLEN



VOORSTELLEN



VOORSTELLEN

Figure 3. Target pictures for *voorstellen* used in the experiment

fact, if *voorstellen*$_1$ 'propose, suggest' and *voorstellen*$_2$ 'introduce, present' are entered as separate verbs in a distinctive collexeme analysis of the corpus data, only the latter verb comes out as significantly attracted to the PD construction (with a distinctiveness score of 11.15), whereas the former verb belongs to the neutral class of verbs.

This brings us into the domain of the second major source of between-corpus variation in subcategorization frequencies discussed in Roland and Jurafsky (2002), viz. *word-sense variation*. One of their examples is the verb *charge*, which they show to be associated with different subcategorization frequencies in the balanced Brown corpus on the one hand and in the Wall Street Journal corpus on the other as a result of the far greater presence in the latter corpus of the business-related senses of *charge* ('accuse' and 'bill'). On the basis of such examples, they posit the *Lemma Argument Probability* hypothesis: the proper locus of probabilistic subcategorization information is the verb sense – the "lemma", in their terms, following Levelt (1989) and others – rather than the orthographic word, i.e. each verb sense contains a vector of probabilistic expectations for its possible argument frames, and the vectors of different senses of one and the same verb may differ in important respects (Roland and Jurafsky 2002: 335–336). Applied to the present topic, *voorstellen*$_1$ 'propose, suggest' and *voorstellen*$_2$ 'introduce, present' are best seen as separate lemmas associated with different lexical alternation biases.

We do not want to suggest that each and every polysemous dative alternating verb can be neatly split up into a number of distinct "lemmas" along these lines – such a suggestion would be quite inconsistent with recent cognitive-linguistic conceptions of the nature of lexical-semantic structure (see, e.g., Geeraerts' 1993: 260 suggestion of a "searchlight" metaphor to replace traditional dictionary conceptions of lexical meaning in terms of lists of neatly separate senses: "[W]ords are searchlights that highlight, upon each application, a particular subfield of their domain of application"). Still, the possibility of distinct senses (i.e., distinct clusters of semantically related uses) with their own alternation biases, as in the case of *voorstellen*, must at least be acknowledged, and, in general, researchers are well-advised to keep an eye open for the possible effects of different (groups of) uses of the investigated verbs when interpreting subcategorization frequency data.

Two more examples may serve to further illustrate this point, viz. the basic verb of transfer of possession *geven* 'give' and its near-synonym *schenken* 'give (as a present)'. *Geven* is strongly attracted to the DO-dative according to the corpus data – it is the second strongest DO-collexeme of the 18 test verbs – but in the experimental data, it belongs to the neutral category (in fact, *geven* displays an above-average preference for the PD construction in the experimental data, which only narrowly misses significance, $p = .061$). The target pictures for *geven* used in the picture description experiment all depict prototypical transfer scenes in which a concrete entity (a book, a cake, or an apple) passes from the domain of control of

human participant A into the domain of control of human participant B (just like in the target picture for *doorgeven* in Figure 1).

It has been observed before, however, that while the use of the neutral GIVE verb to encode such prototypical possessional transfer scenes is typically the first one to come to mind, it is definitely not the most frequently attested use in corpora of natural language. Gilquin (2010) reports on a case study comparing data from the Switchboard corpus and from an online sentence production experiment on English *give*. In the experiment, forty native speakers of American English, on accessing a URL, were shown twenty word stimuli and instructed to type in the very first sentence they could think of including the word. The set of stimuli included two occurrences each of the target verbs *give* and *take*, only the first one of which was taken into account in the analysis, based on the assumption that the first sentence produced for a word reflects cognitive salience better than the second sentence. In the results for *give*, over 43% of the informants' answers instantiated the 'hand' sense of *give*, i.e. its use to encode a prototypical transfer of possession involving a concrete entity, making this the predominant use (the next semantic category reaches only 15.5%). In the Switchboard corpus, by contrast, this intuitively prototypical 'hand' sense only accounts for 11% of all *give* occurrences: it is less frequently attested than several other uses, including the 'communicate' sense as in *to give s.o. one's name*, the 'cause' sense as in *to give s.o. a headache*, the 'allow' sense as in *to give someone the power to do sth* and the delexical sense as in *to give s.o. a kiss* or *a poke in the snoot*, etc. (the labels and examples of these "senses" are Gilquin's). In the data for Dutch *geven* from the CONDIV corpus, the situation is even more extreme. We hand-coded all 504 DO and PD occurrences from one of the six newspapers included in the corpus sample, viz. *De Standaard*, for semantic category (i.e., the nature of the denoted situation), and only 29 of these (i.e., a mere 5.7%) were found to encode a prototypical transfer situation in which a more or less concrete entity changes possession (15 of which are DO-datives and 14 of which are PDs). In comparison, to give but two examples, the use of *geven* in *iemand de kans/de gelegenheid/de mogelijkheid/het recht/de vrijheid/de tijd geven* (*om iets te doen*) 'to give s.o. the chance/opportunity/right/freedom/time (to do sth)' accounts for 64 occurrences, 57 of which are DO-datives, and its combination with a direct object which refers to a property or attribute of the indirect object referent (*een kleur* 'a colour', *een goed/slecht/... gevoel* 'a good/bad/... feeling', *een zekere sfeer/uitstraling/reputatie* 'a particular atmosphere/outlook/reputation', etc.) accounts for 69 occurrences, 62 of which are DO-datives. As shown by these frequencies, both

of these "other" uses display a far greater preference for the DO-dative than the prototypical transfer of possession use of *geven*. The overall alternation bias of *geven* in the data from the CONDIV corpus generalizes over all these different uses of the verb and as such is hardly comparable to the alternation bias computed on the basis of the data from the picture description experiment, which reflect the prototypical use only. In fact, one may wonder whether for semantically highly flexible verbs like *geven*, the overall alternation bias is a very relevant notion at all. In any event, the above findings are compatible with a model that assumes storage of a number of (semi) fixed multi-word patterns such as *iemand een* <Adj> *gevoel geven* 'to give s.o. a <Adj> feeling' and *iemand de gelegenheid/de kans geven* <*om te* Inf> 'to give s.o. the opportunity <to Inf>', with alternation biases of their own.

For a final example, consider the corpus instances of *schenken* in (8) and (9) below. (8) illustrates the basic 'give as a present' sense of *schenken*: there is a concrete transfer of a ring – which qualifies as a prototypical present – from a human giver to a human recipient. Naturally, this is also the sense at stake in the pictures used in the experiment, which depict a cowboy presenting a banana to a thief or a monk giving a book to a doctor, very much like the *doorgeven* scenes in Figure 1. The *schenken* verb also occurs in a number of semi-idiomatic patterns which denote more abstract transfers, however, the most frequent of which is *aandacht schenken aan iets* 'to pay attention to sth', as illustrated in (9). This multi-word pattern accounts for about one fifth of all ditransitive *schenken* instances in the CONDIV corpus.

(8) *Ze is begraven met aan haar vinger de ring, die*
she is buried with on her finger the ring that
Algren haar had geschonken. [NRC]
A. her had given.
'She was buried wearing (on her finger) the ring that Algren had given her.'

(9) *Mannen van de jaren negentig schenken meer*
men of the years ninety give more
en meer aandacht aan hun uiterlijk. [HBL]
and more attention to their appearance
'Men of the nineties are paying more and more attention to their outward appearance.'

Table 5. Observed corpus frequencies for *aandacht schenken* versus *schenken* with other direct objects

	DO	PD	total
aandacht schenken	1	31	32
schenken + all other direct objects	70	69	139
total	71	100	171

Table 5 splits out the observed corpus frequencies of ditransitive *schenken* into *aandacht schenken* on the one hand and *schenken* with all other direct objects (the large majority of which either refers to a concrete object or to a sum of money) on the other. As can be seen from these figures, the predominance of PD instances in the corpus data is to an important extent due to *aandacht schenken*, which occurs virtually exclusively in this construction (the contrast between the two observed distributions is statistically significant, Chi-square = 23.9, $df = 1$, $p < .0001$). It is clear from this brief discussion that *aandacht schenken* is another example of a pattern that is better thought of as a separate lemma in the sense of Roland and Jurafsky (2002), with an alternation bias that differs from the one for "regular" *schenken*. Incidentally, this example also shows that, contrary to what is sometimes assumed (e.g. by Ebeling 2006: 262), such semi-idiomatic multi-word patterns with abstract meanings need not always prefer the *DO* variant: *aandacht schenken* is clearly biased towards the *PD* construction.

6. Some theoretical and methodological implications

The above comparison of corpus and experimental data on the Dutch dative alternation has illustrated the added value to be gained from adopting a collostructional perspective on the phenomenon of alternation bias, i.e. from taking into account the *overall distributions* of the investigated constructions in the dataset in computing the strength of the association between constructions and the individual verbs occurring in them. Lexical alternation biases become all the more relevant if they can be shown to be consistent across different corpora, possibly including corpora of elicited as well as natural language use. We have seen that *verkopen* 'sell', for instance, displays such a robust bias towards the PD construction in Dutch: even in the experimental dataset, which is characterized by a strong overall preference for the PD, *verkopen* still stands out as display-

ing a significantly *above-average* preference for this construction, as shown by the results of a distinctive collexeme analysis. As such, the observed alternation bias of this verb constitutes a solid empirical fact about the dative alternation, which can shed more light on the semantic relation between the constructions involved. As a general methodological guideline, researchers should shy away from labeling individual verbs as DO-biased or PD-biased solely on the basis of their raw observed frequencies in both constructions in a single corpus *without* evaluating these against the overall frequencies of the investigated constructions in the corpus.

The above analysis has also triggered a brief discussion of the relation between lexical alternation bias and *verbal polysemy*. Whereas there is a growing body of literature in which the results from a (distinctive) collexeme analysis of a given construction or pair of constructions provide the starting point for an analysis of the constructional semantics, few of these studies devote any attention at all to issues of verbal polysemy. This lack of attention is somewhat unfortunate, as our comparison of the corpus and experimental results for *voorstellen* 'suggest, introduce, present', for instance, suggests that, in some cases, it may be useful to distinguish two or more "senses" of a verbal lexeme, each with its *own* alternation bias. This proposal is consonant with Roland and Jurafsky's (2002) Lemma Argument Probability hypothesis quoted above (also see Hare, McRae, and Elman 2004 for a relevantly similar proposal involving sense-contingent structural biases).

As was already mentioned in the introduction, Jaeger and Snider (2007) found an effect of lexical verb bias on the strength of structural priming, in that the priming effect is stronger if the prime consists of a verb used in a construction it is biased *against*, which they label *surprisal-sensitive* priming. They point out that these surprisal effects are consistent with *implicit learning* accounts of sentence production, which hold that on every occasion language users process a particular structure, they implicitly learn something about that structure, and that the amount of learning determines the probability of reusing the same structure later on (cf. Chang et al. 2000; Chang 2002; Chang, Dell, and Bock 2006). If language users possess implicit knowledge of the probabilistic distribution of syntactic structures given a certain verb, the processing of an unlikely or dispreferred verb-structure combination will involve a larger amount of implicit learning than the processing of highly frequent verb-structure combinations. Hence, it is to be expected on this account that less expected structures exert stronger priming (see Jaeger and Snider 2007: 26–28 for further elaboration). Jaeger and Snider measure the alternation bias of the verbs included

in their database of dative structures on the basis of the raw number of observed DO and PD occurrences (log[#DO/#PO]), but their approach is not incompatible with a collostructional perspective.

There is another dimension to surprisal-sensitive priming which merits some further discussion, though. Next to the verb-specific effects, Jaeger and Snider (2007) also found evidence for an overall *inverse-frequency* effect, in that, across all verbs, the more frequent structure in their database, viz. the DO construction, was found to prime considerably less strongly than the PD construction (Jaeger and Snider's investigation is based on the set of 1,859 DO and 501 PD instances culled from the Switchboard corpus by Bresnan et al. 2007, see section 3). Such overall inverse-frequency effects have been observed before: Bock's (1986) experimental study, for instance, found a priming effect for the passive structure, but not for the much more frequently produced active structure. In the active-passive alternation, it is of course the passive structure which is consistently the least frequent member of the construction pair, regardless of the exact source of the data (that is, across all verbs, for it is well-known that some verbs do prefer the passive over the active, also see Gries & Stefanowitsch 2004). However, as we have seen, this is different for the dative alternation: in the data from Bernolet's (2008) picture description experiment, unlike in the data from the CONDIV corpus, it is not the PD but the DO construction which is by far the least frequently realized option. We will not present a detailed discussion of the results from the priming part of Bernolet's study in the present paper, but, very briefly, there was indeed an overall inverse-frequency effect, i.e., across all verbs, the DO construction was found to exert stronger priming than the PD construction (see Bernolet 2008: 164–167 for details; also see Bernolet & Hartsuiker 2010). This finding can only be reconciled with the implicit learning account by assuming that speakers' implicit knowledge of the probabilistic distribution of syntactic structures is sensitive to context differences, it would seem, since it is evident from the corpus data cited above that, "in the wild", the DO-dative is definitely *not* the least frequent option. That is, the overall surprisal effect of DO-primes in the experiment is not due to the lower probability of encountering the DO-dative *overall*, but to the lower probability of encountering the DO-dative in the specific linguistic context of the experiment, viz. in an isolated description of a target picture. Additional support comes from the contrast noted by Jaeger and Snider (2007: 40–41) between their own corpus-based results on the priming of complement clauses with or without the complementizer *that* on the one hand and the results of Ferreira's (2003) experimental study on the same topic.

In Jaeger and Snider's spontaneous speech data, the construction *with* an overt complementizer *that* was found to prime more strongly, whereas in Ferreira's production experiment, the construction *without* a complementizer displayed the strongest priming effect. However, in both investigations, it was the *least frequent* member of the construction pair in the database that was found to prime more strongly, thus corroborating the context-sensitivity of the inverse-frequency effect.[17] We leave it to future research to explore the further implications of this position.

7. Conclusion

Our comparison of the frequency data about the Dutch dative alternation reported in Colleman's (2009) corpus-based investigation with the results from Bernolet's (2008) picture description experiment has revealed a striking contrast in overall structural preferences: whereas the DO construction is by far the most frequently realized option in natural language, the experimental data display a distinct overall bias toward the PD construction. We have explored a number of factors contributing to this overall contrast, such as the far greater impact of principles of discourse cohesion in natural language data and, possibly, the greater preference for constructions of maximum perceived clarity and explicitness in production experiments. We have also observed, however, that the alternation biases of the *individual* dative alternating verbs included in both investigations are quite consistent if they are not simply measured in terms of the raw observed frequencies of DO and PD instances in both databases, but in *collostructional* terms, i.e., in a way which evaluates these observed frequencies against the frequencies expected on the basis of the *overall distributions* of the DO and PD constructions in the respective databases. *Verkopen* 'sell', *doorgeven* 'pass, hand (on)' and *uitreiken* 'give out, issue', for instance, have been identified as verbs which display a robust *above-average* preference for the PD constructions. This goes to show that, as a general methodological guideline, researchers should avoid labeling individual verbs as DO-biased or PD-biased solely on the basis of their raw observed frequencies in both constructions in a single corpus of either elicited or spontaneously

17. Also note that, again, it is the most "explicit" construction, i.e. the complement clause with an overt complementizer *that*, which is the most frequently realized option in the production experiment (see the brief discussion of the PD construction's perceived explicitness at the end of section 5.2 above).

produced language corpus, i.e. *without* evaluating these against the overall frequencies of the investigated constructions in that corpus. Finally, we have discussed a couple of verbs for which the results from the two investigations are somewhat less uniform, and we have attributed these different results to polysemy effects and/or to the existence of independently stored multi-word patterns with alternation biases of their own. We believe that, in corpus-based and experimental approaches to argument structure alternations alike, such effects of verbal polysemy merit closer attention than they have as yet received.

References

Bernolet, Sarah
2008 Lexical-syntactic representations in bilingual sentence production, 149–177. Unpublished Ph.D. dissertation, Ghent University.

Bernolet, Sarah and Robert J. Hartsuiker
2010 Does verb bias modulate syntactic priming? *Cognition* 114: 455–461.

Bernolet, Sarah, Robert J. Hartsuiker and Martin J. Pickering
2007 Shared syntactic representations in bilinguals: Evidence for the role of word-order repetition. *Journal of Experimental Psychology: Learning, Memory and Cognition* 33: 931–949.

Bock, J. Kathryn
1986 Syntactic persistence in language production. *Cognitive Psychology* 18: 355–387.

Booij, Geert
2002 Separable complex verbs in Dutch: a case of periphrastic word formation. In: Nicole Dehé, Ray Jackendoff, Andrew Macintyre and Silke Urban (eds.), *Verb-particle explorations*, 21–42. Berlin/New York: Mouton de Gruyter.

Branigan, Holly P., Martin J. Pickering and Alexandra A. Cleland
2000 Syntactic co-ordination in dialogue. *Cognition* 75: B13–B25.

Branigan, Holly P., Martin J. Pickering, Andrew J. Stewart and Janet F. McLean
2000 Syntactic priming in spoken production: Linguistic and temporal interference. *Memory and Cognition* 28: 1297–1302.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina and Harald Baayen
2007 Predicting the dative alternation. In: G. Boume, I. Kraemer and J. Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan and Tatiana Nikitina
2009 The gradience of the dative alternation. In: Linda Uyechi and Lian Hee Wee (eds.), *Reality exploration and discovery: Pattern interaction in language and life*, 161–184. Stanford: CSLI Publications.

Chang, Franklin
2002 Symbolically speaking: A connectionist model of sentence production. *Cognitive Science: A Multidisciplinary Journal* 26: 609–651.
Chang, Franklin, Gary S. Dell and J. Kathryn Bock
2006 Becoming syntactic. *Psychological Review* 113: 234–272.
Chang, Franklin, Gary S. Dell, J. Kathryn Bock and Zenzi M. Griffin
2000 Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Research* 29: 217–229.
Colleman, Timothy
2006 De Nederlandse datiefalternantie. Een constructioneel en corpusgebaseerd onderzoek [Dative alternation in Dutch: A constructionist and corpus-based investigation]. Unpublished Ph.D. dissertation, Ghent University.
Colleman, Timothy
2009 Verb disposition in argument structure alternations: A corpus study of the Dutch dative alternation. *Language Sciences* 31: 593–611.
Colleman, Timothy
2010a Beyond the dative alternation: the semantics of the Dutch *aan*-dative. In: Dylan Glynn and Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches*, 271–303. Berlin/New York: Mouton de Gruyter.
Colleman, Timothy
2010b Lectal variation in Constructional Semantics: Benefactive ditransitives in Dutch. In: Dirk Geeraerts, Gitte Kristiansen and Yves Peirsman (eds.), *Advances in Cognitive Sociolinguistics*, 191–221. Berlin/New York: Mouton de Gruyter.
Colleman, Timothy and Bernard De Clerck
2009 'Caused Motion'? The semantics of the English *to*-dative and the Dutch *aan*-dative. *Cognitive Linguistics* 20: 5–42.
Collins, Peter
1995 The indirect object construction in English: an informational approach. *Linguistics* 33: 35–49.
Corley, Martin and Christophe Scheepers
2002 Syntactic priming in English sentence production: Categorical and latency evidence from an internet-based study. *Psychonomic Bulletin & Review* 9: 126–131.
Ebeling, C.L.
2006 *Semiotaxis. Over theoretische en Nederlandse syntaxis* [Semiotaxis. On theoretical and Dutch syntax]. Amsterdam: Amsterdam University Press.
Ferreira, Victor S.
2003 The persistence of optional complementizer mention: Why saying a "that" is not saying "that" at all. *Journal of Memory and Language* 48, 379–398.

Fillmore, Charles
1977 Topics in lexical semantics. In: R. Cole (ed.), *Current issues in linguistic theory*, 76–138. Bloomington: Indiana University Press.
Geeraerts, Dirk
1993 Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics* 4: 223–272.
Gilquin, Gaëtanelle
2010 Language production: a window to the mind?. In: H. Götzsche (ed.), *Memory, Mind and Language*. Newcastle: Cambridge Scholars Publishing.
Goldberg, Adele E.
1992 The inherent semantics of argument structure: the case of the English ditransitive. *Cognitive Linguistics* 3: 37–74.
Goldberg, Adele E.
1995 *Constructions*: *A Construction Grammar approach to argument Structure*. Chicago: University of Chicago Press.
Goldberg, Adele E.
2006 *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
Gries, Stefan Th.
2003 Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1: 1–27.
Gries, Stefan Th.
2005 Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 34: 365–399.
Gries, Stefan Th., Beate Hampe and Doris Schönefeld
2005 Converging evidence: bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16: 635–676.
Gries, Stefan Th. and Anatol Stefanowitsch
2004 Extending collostructional analysis: a corpus-based perspective on "alternations". *International Journal of Corpus Linguistics* 9: 97–129.
Grondelaers, Stefan, Katrien Deygers, Hilde Van Aken, Vicky Van den Heede and Dirk Speelman
2000 Het CONDIV-corpus geschreven Nederlands [The CONDIV corpus of spoken Dutch]. *Nederlandse Taalkunde* 5: 356–363.
Gropen, Jess, Steven Pinker, Michelle Hollander, Richard Goldberg, and Ronald Wilson
1989 The learnability and acquisition of the dative alternation in English. *Language* 65: 203–257.
Hare, Mary, Ken McRae and Jeffrey L. Elman
2004 Admitting that admitting verb sense into corpus analyses makes sense. *Language and Cognitive Processes* 19: 181–224.

Hartsuiker, Robert J., Sarah Bernolet, Sofie Schoonbaert, Sara Speybroek and Dieter Vanderelst
2008 Syntactic priming persists but the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language* 58: 214–238.

Hartsuiker, Robert J. and H.H.J. Kolk
1998 Syntactic persistence in Dutch. *Language and Speech* 41: 143–184.

Jaeger, T. Florian and Neal Snider
2007 Implicit learning and syntactic persistence: Surprisal and cumulativity. *University of Rochester Working Papers in the Language Sciences* 3: 26–44.

Kirsner, Robert S.
1988 Prepositional versus 'bare' indirect objects in the written Dutch of novels and newspapers. In: Ton J. Broos (ed.), *Papers from the third Interdisciplinary Conference on Netherlandic Studies, held at the University of Michigan, Ann Arbor, 12–14 June 1986*, 279–296. Lanham: University Press of America.

Krifka, Manfred S.
2004 Semantic and pragmatic conditions for the Dative Alternation. *Korean Journal of English Language and Linguistics* 4: 1–32.

Langacker, Ronald W.
1991 *Foundations of Cognitive Grammar. Volume II: Descriptive Application.* Stanford: Stanford University Press.

Levelt, Willem
1989 *Speaking: from intention to articulation*. Cambridge: MIT Press.

Lombardi, Linda and Mary C. Potter
1992 The regeneration of syntax in short-term memory. *Journal of Memory and Language* 31: 713–733.

Manning, Christopher D.
2003 Probabilistic syntax. In: Rens Bod, Jennifer Hay and Stefanie Jannedy (eds.), *Probabilistic Linguistics*, 289–341. Cambridge: MIT Press.

Melinger, Alissa and Christian Dobel
2005 Lexically-driven syntactic priming. *Cognition* 98: B11–B20.

Ozón, Gabriel
2009 Alternating ditransitives in English: A corpus-based study. Unpublished Ph.D dissertation, University College London. <http://eprints.ucl.ac.uk/18714/1/18714.pdf> [last accessed 30/06/2010]

Panther, Klaus-Uwe
1997 Dative alternation from a cognitive perspective. In: Birgit Smieja and Maike Tasch (eds.), *Human Contact through Language and Linguistics*, 107–126. Frankfurt/M: Lang.

Pickering, Martin J. and Holly Branigan
1998 The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39: 633–651.
Roland, Douglas and Daniel Jurafsky
2002 Verb sense and verb subcategorization probabilities. In: Paola Merlo and Suzanne Stevenson (eds.), *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*, 325–345. Amsterdam/Philadelphia: John Benjamins.
Schermer-Vermeer, E.C.
1991 *Substantiële versus formele taalbeschrijving: het indirect object in het Nederlands* [Formal versus substantial language analysis: The indirect object in Dutch]. Amsterdam: Universiteit van Amsterdam.
Schoonbaert, Sofie, Robert J. Hartsuiker and Martin J. Pickering
2007 The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language* 56: 153–171.
Sokolova, Svetlana, Olga Lyashevskaya and Laura A. Janda
this volume The Locative Alternation and Russian 'empty' prefixes: A case study of the verb *gruzit'* 'load'.
Stefanowitsch, Anatol
2006 Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory* 2: 61–77.
Stefanowitsch, Anatol and Stefan Th. Gries
2003 Collostructions: investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8: 209–243.
Stefanowitsch, Anatol and Stefan Th. Gries
2005 Co-varying collexemes. *Corpus Linguistics and Linguistic Theory* 1: 1–45.
Trueswell, John C. and Albert E. Kim
1998 How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure. *Journal of Memory and Language* 39: 101–132.
Van Belle, William and Willy Van Langendonck
1996 The indirect object in Dutch. In: William Van Belle and Willy Van Langendonck (eds.), *The Dative. Volume 1: Descriptive Studies*, 217–250. Amsterdam/Philadelphia: John Benjamins.
Wierzbicka, Anna
1986 The semantics of 'internal dative' in English. *Quaderni di Semantica* 7: 121–35.
Wilson, Michael P. and Susan M. Garnsey
2008 Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language* 60: 368–392.

A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon

*Neal Snider and Inbal Arnon**

Abstract

In this chapter, we address the debate between single-system and dual-system models of language by looking at the processing of multi-word phrases. We present findings that challenge the distinction between 'stored' and 'computed' linguistic forms via two experiments. The first demonstrates parallels in the processing of words and phrases: frequent four-word phrases are processed more quickly than less frequent ones, without any evidence of a frequency threshold. The second experiment shows that idiomatic phrases prime their construction just as well as non-idiomatic phrases, suggesting that they are not stored as unanalyzed wholes, but instead have internal structure. Taken together, the findings undermine the empirical criteria traditionally used to distinguish between 'stored' and 'computed' forms: compositional phrases showed frequency effects, even though such effects are often thought to be a marker of lexical storage, while non-compositional forms (idioms) showed evidence of internal structure, unexpected if they are stored as unanalyzed wholes. The findings show that linguistic structures are processed in qualitatively the same way regardless of where they fall on the frequency and compositionality continua, and highlight the utility of models that deal with all linguistic experience in a qualitatively similar fashion, and allow for experience to influence the learning, representation and processing of all linguistic patterns.

1. Introduction

There has been long-standing tension in the study of language between approaches that assume a clear distinction between the mental lexicon and grammar (dual-system theories, Chomsky 1965, 1995; Fodor 1983; Pinker 1991, 1999; Pinker and Prince 1988; Ullman 2001, 2004) and ones that do not (single-system theories, Bates and MacWhinney 1989; Elman 1991; MacDonald, Pearlmutter and Seidenberg 1994; Rummelhart and

* Both authors contributed equally to this paper.

McClelland 1986; Seidenberg 1994). Dual-system models distinguish between the mental lexicon – an inventory of memorized forms, and the mental grammar – the rules or constraints used to combine the memorized elements. This distinction echoes the one made in many generative models of language (Chomsky 1981; Jackendoff 2002; Kaplan and Bresnan 1982; Pollard and Sag 1994). The mental lexicon is thought to contain the linguistic units that cannot be derived: simple words (e.g. *cat*), morphemes, irregular nouns and verbs, and longer non-compositional phrases like idioms. The most clearly articulated model of this kind is that of Pinker and his colleagues (Pinker 1991; Pinker and Prince 1988; Pinker and Ullman 2002). They propose that the two components of language (lexicon and grammar) are learned differently, involve different cognitive abilities and are governed by different neural substrates (Ullman et al. 2005; Ullman 2001).

In dual-system models, forms created by grammar are distinct from those originating in the lexicon. No such distinction is posited by single-system theories. Instead, all aspects of language depend on one computational system. The same cognitive mechanism processes all linguistic experience, whether a non-compositional lexical item like '*cat*' or a compositional phrase like '*I don't know*'. Word-object mappings and grammatical rules are learned in a qualitatively similar fashion – by abstracting and generalizing from linguistic experience.

A growing number of models implicitly or explicitly take a single-system stand. In connectionist models, the unity of lexicon and grammar is made explicit by using one single network to capture all linguistic experience (e.g., Rumelhart and McClelland 1986; Seidenberg 1994). Exemplar models of language also dispense with the distinction but in a different way, by having linguistic units and categories correspond to clusters of memory traces (Bod 1998, 2006; Goldinger 1996; Johnson 1997; Pierrehumbert 2001). Connectionist and exemplar models differ in several important respects, including the use of symbolic or non-symbolic representations and the implementation of higher level categories like nouns or verbs (see Bybee and McClelland 2005). But both dispense with any clear distinction between 'stored' and 'computed' forms and instead assume that all linguistic experience is learned, processed and used in a similar fashion.

These models are closely related to what are often labeled usage-based approaches to language where grammatical knowledge emerges from linguistic experience (Bybee 1998, 2006; Goldberg 2006; Barlow and Kemmer 2000; Langacker 1986, 1987; Tomasello 2003). The lexicon is not 'reserved' for atomic elements. There is no a priori limit on the size of the units that

are stored; as long as they can be attended to and remembered, they can be of varying length (word, two-word, multi-word phrase) and levels of abstraction: from single words, through partially realized constructions to fully abstracted ones (*give*, *give me a break*, *give NP a break*, *give NP NP*, *V NP NP*).

1.1. The representational status of multi-word phrases

The contrast between dual-system and single-system approaches has been studied primarily in the domain of morphological representation and processing (Rumelhart and McClelland 1986; Pinker and Prince 1988). But given their diverging assumptions about language, these approaches make different predictions about many aspects of language use. In this chapter, we contrast the two approaches by looking at a relatively less-studied domain: the processing of larger units of language – multi-word phrases. We use this expression to refer to multi-word sequences that are syntactic constituents (e.g. *don't have to worry*, but not *in the middle of the*).

The ways the two approaches handle words are clearly articulated in existing models (e.g., Rumelhart and McClelland 1986; Pinker 1991). Their predictions about larger units are not clearly stated in any existing model but can be extrapolated from their general assumptions about language. Just as they differentiate between regular and irregular morphological forms, dual-system models maintain a distinction between compositional phrases (like *don't have to worry*) and non-compositional ones (as in idioms like *gave the surfer the creeps*). Compositional phrases are generated by the grammar while non-compositional ones originate in the lexicon and are stored together with their idiosyncratic syntactic and semantic features. Idioms should have the characteristics of stored forms while compositional phrases should not. Because compositional phrases can be derived in a predictable way, there is no need to store them in the lexicon. In fact, given the goal of minimizing storage (e.g., Pinker, 1991), compositional multi-word phrases would seem unlikely candidates for storage in the lexicon.

In contrast, single-system models do not posit such a distinction. Multi-word phrases, whether compositional or not, should be like any other linguistic pattern. Every encounter with a phrase is predicted to add to its representation and influence future processing. Compositional and non-compositional phrases should be impacted by the same factors (e.g., frequency) that impact the processing of both bare and regularly inflected words. The two kinds of phrases should also be processed in a qualitatively similar fashion: to the extent that compositional and non-compositional

phrases share structural and lexical features, they should be processed in the same way. For example, hearing a compositional dative phrase like *give the man a hammer* makes one more likely to re-use the double object construction in future dative uses (e.g. Bock 1986; we expand on syntactic priming in section 3). If non-compositional phrases involve similar syntactic processes, then hearing a non-compositional phrase like *give the man a lift*, which has the same head verb and can also alternate, should also increase the likelihood of using the double object construction.

The extent to which compositional multi-word phrases are part of the mental lexicon, and the extent to which their status can be distinguished from that of non-compositional phrases, has an important role in evaluating models of language. In this chapter, we show that is it hard to differentiate compositional and non-compositional phrases empirically, and in doing this; we argue against the distinction (posited in dual-system models) between 'stored' and 'computed' forms more generally.

We do this in two ways: first, we demonstrate that compositional phrases (like *don't have to worry*) exhibit phrase-frequency effects similar to those found for words. Such a finding shows that speakers are sensitive to the frequency of a range of units (including ones that are 'computed' under dual-system models). It also undermines the empirical distinction between stored and computed forms: generated forms display frequency effects thought to be a mark of lexical storage under dual-system models (e.g. Ullman and Wellensky 2005). Second, we show that idiomatic and non-idiomatic datives prime their syntactic construction to a similar degree. Such a finding again blurs the distinction between compositional and non-compositional phrases: forms that are considered to be 'stored' (e.g. idioms) maintain internal structure and activate their constructions just like 'computed' forms. Together, these findings reveal similarities between 'stored' and 'computed' forms and undermine the possibility of coming up with empirical criteria to distinguish the two.

In Section 2 we report on studies showing that speakers are sensitive to the frequency of four-word compositional phrases. In Section 3 we present a novel study showing that idiomatic and non-idiomatic datives prime to a similar degree. In Section 4 we discuss these results in light of the contrast between dual- and single-system models of language.

2. Comprehenders are sensitive to the frequency of compositional phrases

In this section, we report a series of experiments published in Arnon and Snider (2010) showing that people process more frequent 4-word phrases

faster than less frequent ones. This effect occurs across the frequency continuum, with no evidence of a threshold or cutoff. But first we take up the role frequency plays in single-system and dual-system models.

2.1. Frequency effects and mental representation

Frequency plays a very different role in single and dual-system models. In single-system models, frequency – as an approximation of experience – plays a central role in the emergence and entrenchment of linguistic units. The more often a pattern is experienced, the easier it becomes to access and use (Bybee 2006; Bod et al. 2003; Bybee and Hopper 2001). Single-system models differ in the specific mechanisms they use to explain the processing advantage of more frequent forms (by impacting the weights in a connectionist network; by lowering the threshold of activation in spreading activation networks; or by enhancing the activation of a memory trace in exemplar models). But they share a common belief that frequency effects inform us about the units that speakers attend to, and predict that frequency effects should be found for all linguistic units: simple and complex.

Frequency effects are viewed differently in dual-system models. The role of frequency in language representation and use is rarely discussed explicitly in these models (e.g. Pinker 1999). This absence echoes the traditional view in generative linguistics that frequency effects are irrelevant to the study of language because they reflect real-life probabilities or performance issues that are separate from, and immaterial to, linguistic knowledge (Chomsky 1957; recently re-argued for by Newmeyer 2003). In some models frequency effects are relegated to the mental lexicon (Ullman and Wellensky 2005). This allows them to account for the widespread frequency effects found in word production and comprehension (see Monsell 1991) while maintaining that 'stored' elements should exhibit frequency effects but 'computed' elements should not.

2.2. Lessons from morphology

Frequency effects have been used to contrast single-system and dual-system models of regular and irregular inflected forms (e.g. *walked* vs. *felt*). Dual-system models predict that irregular forms will be stored in the mental lexicon while regular forms will be generated by the grammar (Marcus et al. 1992; Pinker 1991, 1999; Pinker and Prince 1991; Pinker and Ullman 2002; Ullman et al. 1997). Single-system models predict that all forms will be represented by the same associative memory mechanism (Rumelhart and McClelland 1986; Plunkett and Marchman 1991, 1993; Marchman 1993).

If regularly inflected forms cannot be accessed as whole words, then the base form (e.g., *walk*) should be activated every time an inflected form is encountered. Access speed should reflect the frequency of the base in all its various inflections (e.g. walks, walking, etc.). If a whole-word representation is available, then the frequency of the inflected form should also affect access speed. Finding that the frequency of the inflected form is predictive of processing time suggests a whole-form representation is available, as argued by single-system, but not dual-system, models. Indeed, the frequency of the inflected form itself (*walked*) predicts processing latencies when the frequency of the base form (*walk*) and the inflectional morphemes (*-ed*) is controlled for (e.g., Alegre and Gordon 1999; Baayen et al. 1997; Taft 1979).

A similar whole-form frequency manipulation has been extended to the study of phrases in child language (Bannard and Matthews 2008). Two and three-year-olds are faster and more accurate at repeating higher frequency phrases compared to lower frequency ones when part frequency is controlled for (e.g. *a drink of tea* vs. *a drink of milk*). Children are sensitive to phrase-frequency. This in turn suggests that they represent whole phrases at some level, just as in whole-word representation of regularly inflected words.

2.3. Phrase-frequency effects

In a series of studies we used a manipulation similar to that used by Bannard and Matthews (2008) to look at the processing of compositional phrases in adults (Arnon and Snider 2010). We wanted to see (a) whether adults are sensitive to phrase-frequency, and (b) whether this holds not only for very frequent phrases, but whenever a higher-frequency phrase is compared to a lower-frequency one. Language-users should be sensitive to phrase-frequency according to single-system, but not dual-system, models. We undertook the latter analysis to test the predictions of a slightly modified dual-system model that allowed very frequent phrases to be stored in the lexicon. Very frequent forms have privileged status also in specific usage-based models (e.g., Goldberg 2006). Therefore, asking whether there is a threshold for phrase-frequency effects has implications for those models as well (see Arnon and Snider 2010 for a further discussion).

2.3.1. Previous research

Many studies have shown that two-word (bigram) frequency affects processing: words are faster to process (McDonald and Shilcock 2005; Reali and

Christiansen 2007) and shorter to produce (Bell et al. 2003, 2009; Gregory et al. 2004; Jurafsky et al. 2001) when they appear as part of a more frequent bigram. People keep track of co-occurrence patterns for single words, but capturing such relations doesn't require any representation beyond the single word. Few studies have looked beyond the bigram, and most of those have focused on the processing of highly frequent phrases. For instance, Bybee and Scheibman (1999) found that *don't* was phonetically reduced in the frequently recurring phrase *I don't know*. Bell et al. (2003) likewise found that the ten most frequent words in English are phonetically reduced when they are more predictable given the previous and following word. Bannard and Matthews (2008) showed that children are sensitive to phrase-frequency but their frequent items were also taken from the top third of the frequency range.

A few other studies have looked at frequency beyond the bigram for a broader frequency range. Levy and Jaeger (2007) found an effect of predictability, given the previous two words, on relativizer omission in English relative clauses. Speakers were more likely to omit the relativizer when it was more predictable given the last one, two, and three words of the pre-relative clause utterance, but because they do not report the independent effect of each string size (this was not the goal of their paper), we cannot know whether their results show an effect of three-word frequency when bigram and unigram frequency are controlled for. Underwood, Schmitt, and Galpin (2004) used eye-tracking to look at participants' eye-movements while reading formulaic sequences of up to six words (e.g., *as a matter of fact*). They found fewer fixations when words appeared in formulaic sequences, which they interpreted as evidence that people represent the sequences as a whole. But since they did not control for the frequency of the substrings either, or for the plausibility of each phrase (plausibility isn't controlled also in Bannard & Matthews, 2008), it is hard to know how to interpret their results.

These effects provide limited evidence that adults are sensitive to the frequency of compositional phrases. We need more evidence from adults, with part frequency and plausibility controlled for, and from phrases across the frequency continuum.

2.3.2. Our findings

We conducted two reaction times studies where we compared processing latencies for pairs of compositional four-word phrases that differed in phrase frequency (the frequency of the four-word phrase) but were matched

for part frequency (unigram, bigram, and trigram frequency), and for plausibility relative to the event they describe (e.g. *don't have to worry* vs. *don't have to wait*). We measured processing latencies using a phrasal decision task. People saw four-word phrases and had to judge whether they were possible in English. We used this task for two reasons. First, lexical decision tasks are often used in the study of morphologically complex words (e.g., Baayen et al. 1997). Since we are using a similar frequency manipulation (varying the frequency of the whole form vs. the parts), we wanted to use a similar task. Second, the task allows for the presentation of the phrase as a whole and encourages participants to attend to each phrase as a unit. We controlled for the frequency of the sub-strings by comparing phrases that differed only on the final word, and by controlling for the final word, the bigram, and the trigram, both in the item selection and in the statistical analysis of the results. We also controlled for the plausibility of the events depicted by the phrases using a norming study.

The two experiments together looked at phrases in three frequency bins, in order to test the effect of frequency across the spectrum. The High frequency bin compared phrases that occurred above ten times per million in the corpus, with those that occurred below ten per million. The Mid frequency bin compared phrases between five and ten per million with those below 5 per million. The Low frequency bin compared phrases between one and five per million with those below 1 per million. The items were constructed using a 20-million word corpus that consisted of the Switchboard (Godfrey, Holliman, and McDaniel 1992) and Fisher (11,699 recorded telephone conversations in American English, 18 million words; Cieri, Miller, and Walker 2004) corpora. In each bin, the high and the low variant differed in phrase-frequency but were matched on all other measures, including plausibility. Table 1 gives example items from the different bins, together with their phrase-frequency.

49 Stanford students were paid to complete the two experiments. All were native English speakers. Each participant saw one four-word phrase on the screen at a time and had to decide (as quickly as possible) whether they were possible sequences in English while their response time was measured. The experiments had an equal number of possible and impossible sequences (fillers). During a practice phase, *I saw the man* was given as an example of a possible sequence, and *I saw man the* and *jump during the pool* as impossible sequences.

We analyzed the data using mixed-model linear regression. As predicted, higher-frequency phrases were decided on faster than lower frequency phrases in all three bins. We then took the responses from all three bins

Table 1. Mean frequency (per million words) and example items in the three bins (N = number of items).

High bin ($N = 16$) (High: 19.48, Low: 3.61)		Mid bin ($N = 12$) (High: 9.75, Low: 0.75)		Low bin ($N = 17$) (High: 3.5, Low: 0.2)	
Don't have to worry	15.3	It takes a lot	7.35	Don't have any money	2.35
Don't have to wait	1.5	It takes a little	1.25	Don't have any place	0.2
I don't know why	35.5	all over the country	9.55	I want to sit	3.6
I don't know who	7.0	all over the house	0.85	I want to say	0.2

and conducted a meta-analysis of the reaction times that compared how well a continuous measure of frequency fit the data compared with a categorical one (high vs. low, calculated from the best-fitting breakpoint of frequency). We found a continuous effect of frequency on reaction times across the continuum, and this was a better fit than the categorical measure. Figure 1 shows the model fit with average log reaction times in 6 frequency bins. The fit line shows that the more frequent the phrase, the faster participants respond to it. The fit is derived from a regression

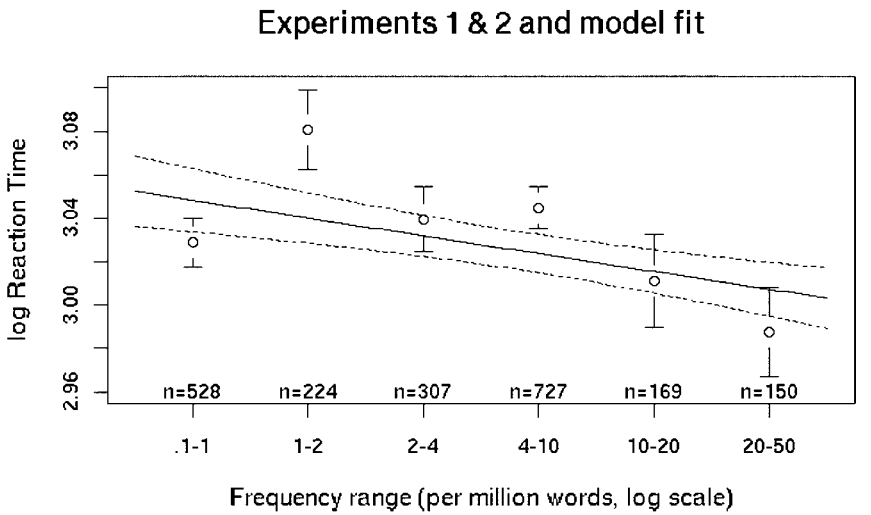


Figure 1. Model fit for reaction times to all phrases. Log reaction time by sequence frequency bin (log scale). Circles represent the means for each bin, with 95% confidence intervals. The fit line is derived from a regression model with a continuous measure of frequency and all control covariates and also includes 95% confidence intervals

model with a continuous measure of frequency and all control covariates, so it reflects the effect of 4-gram frequency beyond the frequencies of the subparts of the phrases.

2.4. Discussion

Our findings show that higher frequency phrases are processed faster across the frequency range. The meta-analysis revealed a direct relation between frequency of occurrence and processing latency: the more often a phrase had been experienced, the faster it was processed.

The current findings are hard to accommodate within a strong dual-system model like the words-and-rules model (Pinker 1999) where frequency effects are taken as a marker of lexical storage. Compositional units (regular words or compositional phrases) are not expected to display whole-form frequency effects because they are not stored as such. One way to explain these effects is to allow for compositional forms to be stored. This is the solution adopted by Ullman and Wallenski (2005) to account for the frequency effects found for regularly-inflected words. Our current findings would require this model to extend the lexicon dramatically to include many (if not all) compositional phrases. It is no longer clear what, if anything remains outside the lexicon, thus undermining the distinction between the mental lexicon and grammar that these models depend on.

The distinction could also be maintained if compositional forms could be both 'stored' and 'generated'. Sometimes phrases would be stored (resulting in frequency effects) and other times generated. This solution runs into an equally difficult problem. It is not clear when speakers use each type of phrase or how this can be tested empirically. The results are also not easy to accommodate within a 'weak' dual-system model that posits a unique status for very frequent forms, for there was no indication of a clear difference between very frequent and low frequency phrases. Frequency effects were found across the continuum. Using a frequency threshold as a determiner of storage is problematic because speakers cannot know a priori which phrases will become frequent enough to merit storage. Whatever information is maintained for very frequent phrases must have once been registered for all phrases. This information could be discarded at later stages of learning, but this seems improbable.

The results are most compatible, however, with single-system models where frequency is expected to affect all linguistic forms in a similar way. Compositional phrases showed whole-form frequency effects like those displayed by simple and inflected words.

In the next section, we look more closely at the postulated distinction between compositional and non-compositional forms from another perspective: We ask whether idioms, often thought to be stored in the lexicon in dual-system models, are processed differently from compositional phrases. We do this by conducting a syntactic priming experiment to see if, and to what degree, idiomatic and non-idiomatic phrases activate the syntactic structure they occur in.

3. Priming from idiomatic and non-idiomatic datives

In this section, we report an experiment that shows that idiomatic datives prime their syntactic structure (make it more likely to be repeated) just as well as non-idioms do. But first, we take up the status of idioms and non-idioms in dual-system and single-system models.

3.1. Introduction

The processing of idioms further blurs the distinction between 'stored' and 'computed' material. Idioms are often seen as prototypical candidates for 'storage' in dual-system models because of their non-compositional character (Pinker 1999; Jackendoff 1995). Take a prototypical example like *kick the bucket*: the meaning of this phrase (at least synchronically) is radically different from what would be expected given typical uses of *kick* and *bucket*. The meaning 'die' arises from the idiosyncratic interpretation of this particular combination of lexical items. More generally, idioms cannot be transparently derived from their parts. To deal with this, idioms are assumed to be stored in the lexicon as a single entry that contains their special lexical, semantic and syntactic features (e.g. Jackendoff 1997). Such a view suggests that idiomatic and non-idiomatic phrases are generated, processed, and retrieved differently. Since idioms are stored together with their structural information, they should not undergo the same syntactic processes as non-idiomatic phrases. Much of the research on idioms assumes a dual-system view in which idioms are stored in the lexicon, with the debate centering on the degree to which they have internal structure (how compositional they are, e.g., Nunberg, Sag and Wasow 1994) and on the kind and amount of syntactic information represented in their lexical entries (e.g., Cutting and Bock 1997; Sprenger et al. 2006).

In an influential article, Nunberg, Sag, and Wasow (1994) assumed a dual-system model but argued that many idioms are compositional, thereby

limiting the number of idiomatic expressions that need to be stored. They pointed out that many structures with highly metaphorical or idiomatic meaning derive that meaning from a metaphorical sense of the words involved, which is then computed by regular syntactic processes. In their account, a phrase like *pull strings* derives its idiomatic meaning from a metaphorical use of *strings*, meaning something like "connections", but the structure is otherwise compositional. On the other hand, *kick the bucket* is an idiomatic construction in their theory since there is no metaphorical sense of *kick* or *bucket* that yields the meaning 'die'. This view clearly predicts differences between the syntactic processes involved in idiomatic and non-idiomatic constructions.

Single-system models acknowledge that while there is a lot of idiosyncrasy peculiar to idioms, they share a lot of structural similarities with other syntactically compositional structures. They therefore blur the distinction between storage and computation by allowing for redundancy in linguistic representation. Goldberg (2006) pointed out that regular linguistic patterns are often instantiated by exemplars that are highly idiomatic. For example, an idiom like *go kicking and screaming* <*path*> is structurally an exemplar of the general pattern *go*VP*ing*. Expressions like *the bigger they come, the harder they fall*, and *the more the merrier*, are instances of the more general *the* X*er the* Y*er* pattern, even though they have their own special features as well. Idioms are tokens of more general (and regular) patterns in addition to being tokens of their own more specific patterns. Given this analysis, and assuming that all linguistic material is processed by a similar mechanism, a single-system account would predict (1) that idioms should maintain links with the more general (and regular) patterns they are instances off, and (2) that, in doing so, they will be similar to non-idiomatic expressions.

3.2. How to distinguish idioms from non-idioms?

Dual-system models treat idioms and non-idioms as qualitatively different entities, but such a distinction is not easy to operationalize. One key issue in the study of idioms is how they should be defined: what makes something an idiom? Such a definition is hard to come by since idiomatic phrases seem to fall on a continuum of compositionality (how transparent their meaning is given their parts) and flexibility (how flexible they are in terms of the lexical items used, number, tense, etc.), with both factors contributing to their perceived status as an idiom (Jackendoff 1997; Nunberg, Sag, & Wasow 1994; Wulff 2008).

Empirically, many findings highlight the inherent complexity in classifying idioms. Idiomatic expressions don't fall neatly into compositional and non-compositional. Idioms differ in their degree of compositionality. Some idioms get their idiomatic meaning more from individual words than others. For example, changing a phrase like *kick the bucket* to *kick the pail* doesn't evoke the idiomatic meaning of 'die'. But a similar change from *pop the question* to *pop the request* still retains some of the idiomatic meaning of 'propose marriage' (Gibbs & Nayak 1989). That is, even phrases that seem highly non-compositional show some degree of reliance on their parts for meaning. Idioms also fall on a continuum with regard to their flexibility: whether the idiomatic meaning is retained in different syntactic constructions (e.g., passivization), morphological realizations (e.g. change of person, number, etc.), and lexical substitutions. Flexibility also seems to be a matter of degree, and is affected by various factors (Wulffe, 2008). Idiomatic phrases can be more flexible in one dimension than another. For example, the idiomatic phrase *throw in the towel* cannot be passivized (*the towel was thrown in* does not mean 'quit') but the idiomatic meaning is retained when the verb is substituted with *toss*. In sum, neither flexibility nor compositionality provide a clear-cut way to distinguish between idioms and non-idioms; idiomaticity seems to a gradient notion and not a categorical one.

In light of these findings (and because resolving this quandary is beyond the scope of the current chapter), we adopt a working definition of idiomatic phrases based on semantic compositionality taken from Nunberg et al. (1994), also adopted by Konopka and Bock (2009): An utterance is idiomatic to the degree that its meaning is not predictable from any regular sense of the words involved. Importantly, our definition of an idiom is gradient, not categorical. Our claims should be understood accordingly: in this chapter, we are interested in investigating parallels in the processing of phrases differing in their degree of idiomaticity.

Several results reveal such parallels between the processing of idiomatic and non-idiomatic phrases. Comprehension and production findings show that literal word meanings are activated during idiom processing (Cacciari and Tabossi 1988; Cutting and Bock 1997). Sprenger et al. (2006) showed that idioms can prime and be primed by words that appear in them (e.g. *hit the road* primes *road*), suggesting that like compositional phrases, they have internal structure. Konopka and Bock (2009) showed that idiomatic and non-idiomatic phrasal verbs (e.g. *pull off a robbery*) can prime particle placement (whether the particle appears before or after the direct object) in non-idiomatic phrases that have no lexical overlap (e.g. *knocked over*

the vase can prime *pull off a robbery*, see section 3.3). Using acceptability judgments of familiar and invented idioms, Tabossi, Wolf, and Koterle (2009) suggested that the syntax of idioms is governed by syntactic and pragmatic principles qualitatively similar to those that govern non-idiomatic language.

In this section, we add to these studies by providing further evidence that idiomatic and non-idiomatic datives prime their syntactic construction to a similar degree. Such a finding (1) enhances the idea that idioms have internal syntactic structure and (2) undermines the possibility of distinguishing empirically between idiomatic and non-idiomatic forms, a distinction predicted by dual-system, but not single-system, models.

3.3. Using syntactic priming to compare idiomatic and non-idiomatic phrases

Several methodologies have been applied to idiom processing. Syntactic priming is particularly interesting because it offers insight into the representational similarity of structures. In syntactic priming, syntactic structures are re-used by speakers, as in the following dialogue from the Switchboard corpus (Godfrey 1992):

(1) I don't feel we should *loan them money*...

I wish our leaders were really seeking the Lord on these things, and if we feel led to *give a country money* to help them, fine...

The speaker first chooses the Double Object (DO) dative construction *loan them money*, even though the Prepositional Object (PO) construction is possible (*loan money to them*). Later, when the speaker produces another dative, they again choose the DO alternate, possibly because of priming from the previously produced dative. Priming was first commented on by sociolinguists (Sankoff & LaBerge 1978; Poplack 1980; Weiner & Labov 1983; Estival 1985), but experimental psychologists (Bock 1986; Pickering & Branigan 1998) have since argued strongly for its role in illuminating representations in language processing. In the psychological literature, priming is seen as a general process (i.e. occurring in both production and comprehension) where the processing of a stimulus (the 'target') is facilitated if a similar stimulus (the 'prime') has just been processed. This facilitation is greater the more similar the prime and the target, and in fact only occurs if they are similar along some cognitive dimension. As Branigan *et al* (1995) argue, this is why priming can illuminate the mental representation of linguistic knowledge, because if people's behavior is sensitive

to this similarity, it indicates that the two structures share a cognitive representation on some dimension. Thus, by exploring the dimensions of similarity experimentally between primes, one may gain insight into the mental representations of the relevant stimuli. The dependence of priming on similarity is important: an utterance should prime a construction (i.e. make it more likely to be repeated, or more easily comprehended) only if it is perceived to be an instance of that construction. Priming thus becomes an important diagnostic for determining whether idioms are instances of the more abstract (and regular) constructions they appear in.

Single-system and dual-system models make different predictions about the priming of idioms. 'Strong' dual-system models (Jackendoff 1997) argue that idioms are stored separately from superficially similar structures with similar word orders. They would predict that idioms should not prime superficially similar structures. For example, a compositional phrase like *give the child some food* is an instance of the double-object dative pattern. However, an idiom like *give the child a lift* is stored separately, and is not a token of the double-object dative pattern. An idiom should therefore not be able to prime a compositional structure like the double object dative because it is not structurally similar. 'Weaker' dual-system models may allow idioms to have internal syntactic structure (Chang, Dell, and Bock 2006), but would probably predict that idioms would prime less than non-idiomatic phases because the link to the construction is weaker. In single-system models, idiomatic and non-idiomatic phrases are represented in the same way (Goldberg 2006). Idioms therefore have internal structure: to the extent that two structures share features like lexical items, argument order, and syntactic construction, they should prime one another. Therefore an idiomatic dative should prime a non-idiomatic one.

3.4. Previous work with priming

The first experiment to examine semantic compositionality and priming was Konopka and Bock (2009). They did a production priming study of the verb-particle alternation where a particle precedes or follows the object NP (e.g., *A celebrity threw in the first ball.* vs. *A celebrity threw the first ball in.*). The task was to repeat a sentence that had been presented rapidly, one word at a time, in the center of a screen. People sometimes mis-repeat the target sentence and use the other alternant instead of the original. Konopka and Bock measured whether this tendency to mis-remember increased when the other alternant was primed by appearing in the previous sentence, and indeed they found priming in this alternation. They went on

to manipulate the idiomaticity of the prime sentence, as determined by a norming task where idiomaticity was defined as the extent to which the meaning of the sentence deviated from that expected given the "dictionary definitions" of the words in the sentence. In this way, they took into account the points of Nunberg, Sag, and Wasow (1994), by defining idioms as *constructions* with idiomatic meaning, not structures that derive their metaphorical or idiomatic interpretation from the metaphorical or idiomatic senses of the words. They found that idiomaticity had no effect on priming: idiomatic verb-particle constructions (*The teenager shot off his mouth.*) were just as likely to be repeated as non-idiomatic ones were (*Judy snapped on her earrings.*). They also looked at the effect of flexibility (whether the structure can appear in the second alternant) on how likely the structure is to be repeated. This is related to the hypothesis that flexibility is correlated with semantic compositionality and hence the idiomaticity of the construction (Jackendoff 1997; Nunberg *et al.* 1994). They manipulated flexibility independently (along with idiomaticity) and found a main effect of flexibility in that frozen structures (e.g., *The crooked salesman couldn't take the customer in*) were less likely to be repeated than flexible structures (*The graduating senior sent his application in*), but found no interaction with idiomaticity. Their findings showed that idiomaticity does not affect whether a structure primes, suggesting similarity in the syntactic processes associated with idiomatic and non-idiomatic structures.

Given that this is the only experiment to date that has examined compositionality and production priming, we wanted to look more closely at the effect of idiomaticity (as measured by semantic compositionality) on priming. We conducted another experiment using a different methodology and a different syntactic alternation. We wanted to use a method closer to natural production where participants have more freedom in what they produce, so we chose a sentence completion task, where participants complete sentence fragments. And we used the dative alternation where the double object (DO) structure (*The mother gave the hungry baby some food*) alternates with the prepositional object structure (*The mother gave some food to the hungry baby*), because this construction lends itself well to completion tasks (Pickering and Branigan 1998).

Strong dual-system models predict that idioms should not prime the structure that they occur in, or at least prime it less, because they are represented in a fundamentally different way from a superficially similar compositional structure. Single-system models predict that idioms should prime their structure just as well as non-idioms do, if they share similarities like lexical items and argument order.

3.5. Syntactic Priming Experiment

We did a production priming experiment of the dative alternation that manipulated the idiomaticity of the prime.

3.5.1. Method

3.5.1.1. Participants

Thirty-five students (mean age 20 years) from the University of Rochester participated in the study. All were native English speakers and were paid $7.50 in return for their participation.

3.5.1.2. Procedure

We used a sentence completion task (Pickering and Branigan 1998) to assess production priming. Participants saw partial sentences (one at a time) on a screen, and were instructed to complete them in the most sensible way, succinctly, without using pronouns, and to type the entire sentence (not just their additional material) into the input box. Participants were told that if a word or phrase appeared in parentheses after the fragment, they should use that material in the completed sentence. This ensured that the desired recipient and theme were used. Participants saw sequences of prime and target sentences, with fillers appearing between each prime-target sequence. Prime sentences contained enough material to force participants to complete them with the desired alternation: for the DO condition, the sentence fragment included the recipient (e.g. "The mother gave the hungry baby (some food)"), and for the PO condition, the sentence fragment included the theme and the preposition '*to*' (e.g. "The mother gave some food to (the hungry baby)"). The target sentence fragment contained only a subject NP and a dative verb (e.g. "The flight attendant gave"), and could be completed with either alternative. The experiment was conducted using Linger (developed by Douglas Rhode, http://tedlab.mit.edu/~dr/Linger).

3.5.1.3. Materials

The experiment contained 24 items, with each item appearing in two conditions that varied in prime construction (Double Object vs. Prepositional Object). Idiomaticity was manipulated between items (based on a measure of idiomaticity derived via the norming experiment described in the next section). Our choice of theme determined whether the utterance was idiomatic or not. The two item variants (DO or PO) were followed by the same sentence fragment to elicit the target. A sample item is illustrated in Table 2:

Table 2. Example materials for the priming experiment

Prime:		
Higher idiomaticity	DO:	The lifeguard gave the surfer (the creeps)
	PO:	The lifeguard gave the creeps to (the surfer)
Lower idiomaticity	DO:	The mother gave the hungry baby (some food)
	PO:	The mother gave some food to (the hungry baby)
Target:		The flight attendant gave

More and less idiomatic datives were extracted from the British National Corpus (BNC, the automatically parsed version of Roland et al. 2007). The BNC corpus was used because of its size and availability in at least an automatically parsed form, in order to have a sufficient number of idioms and their frequencies. We first extracted all dative sentences where the verb-theme combinations were of sufficient frequency (over 10 times in the BNC corpus). We selected more and less idiomatic verb-theme combinations that we then normed for idiomaticity (using the definition of Nunberg, Sag, and Wasow 1994 discussed above) and alternation bias (how likely they are to appear in either construction). The norming was done by American English speakers to fit the language experience of the American participants who participated in the priming experiment. We provide more details about this in the next section. The vast majority of dative idioms in the corpus involved the verb '*give*', so much so that 12 idiomatic items could not be constructed with a reasonable variety of verbs. We therefore decided to use only *give* in all the primes and targets (but obviously not the fillers). Given that '*give*' is used in 80% of datives in spoken language (Bresnan et al. 2007), we assume that participants would not notice the high frequency of '*give*' in the experiment. As a further precaution, we presented only half the items, counterbalanced, to each subject, so they would not see too many tokens of '*give*'. We also selected the items with respect to how they were scored on norming tasks measured idiomaticity and flexibility (whether there was a strong bias towards PO and DO), as described below. The materials were presented by Linger in 8 randomized lists using a latin-square design, and each participant saw only one of the two variants of each item.

3.5.2. Norms

3.5.2.1. Idiomaticity

The idiomaticity of each variant was determined using a rating task performed over the web (on Amazon Mechanical Turk, www.mturk.com).

Participants were asked to judge the idiomaticity of each item. Idiomaticity was defined just as in Konopka and Bock: how predictable the meaning of the sentence is given the "dictionary definitions" of the words involved. Participants rated idiomaticity on a 1–7 scale, with 7 being highly idiomatic and 1 being highly non-idiomatic. All items were presented in the DO alternant. 10 ratings were collected for each stimulus (2 conditions per item, for 40 stimuli). Because participants on Mechanical Turk do not have to complete the entire experiment (and often do not), 40 people participated, with the only restriction on participation being that their IP address be from the United States (so that they will resemble the test population). Each item consisted of one task page in the Mechanical Turk interface, with a filler occurring before each experimental item. Participants were paid \$0.02 for each stimulus completed.

We analyzed the norming task, and found that there was a significant difference in idiomaticity judgments between the "idiom" and "non-idiom" items ($t(31) = 24$, $p < .001$): "idiom" items had a mean of 4.2 (range 3.7–4.8), and "non-idiom" items a mean of 2.2 (range 1.1–2.7).

3.5.2.2. Alternation bias

We performed a further norming experiment to determine the bias of each of the 24 items towards the DO or PO construction. We did this for two reasons: First, Konopka and Bock showed that idiomaticity and flexibility were independent factors, and we wanted to manipulate idiomaticity independently of flexibility, so we ensured that all items were flexible, they could occur plausibly in both alternations. Second, since Konopka and Bock found an effect of alternation bias on priming such that structures that do not alternate also prime less, we wanted to be able to add item-bias as another factor in our analysis.

Participants were asked to compare the acceptability of the PO alternant versus the DO one using magnitude estimation (Bard et al. 1996). One alternant was set as a baseline (with a score of 100), and participants were asked to judge how many times more or less acceptable the other alternant was by comparison. Which alternant was presented as the baseline was randomized, and only one condition was presented per item per participant, with the condition selected at random. Each item consisted of one task page in the Mechanical Turk interface, with a filler occurring before each experimental item. Some fillers included what we thought would be non-alternating datives (extremely biased towards PO or DO) as a comparison (e.g. "*The captain gave the old sailor the willies.*"). Participants were paid \$0.02 for each stimulus completed. 24 judgments were collected

per item (for an average of 12 per stimulus), and 133 people participated, restricted to United States IP addresses. The norming results confirmed that all items were indeed variable (all experimental stimuli fell within 2 standard deviations of the mean log odds with respect to their alternation bias).

3.5.3. Fillers

Each experimental item (prime-target pair) was separated by at least 2 fillers, with a total of 42 fillers, and the first 4 were part of a practice block. Half of the fillers used intransitive verbs, one quarter used monotransitive verbs in the simple past tense in order to elicit the active voice, and one quarter used monotransitive verbs in the passive participle and with the preposition '*by*' in order to elicit a passive. Active and passive fillers were presented in order to distract from the dative alternations being elicited in the main experiment and to mask the true object of study.

3.6. Results

Each response from the participants on the experimental stimuli (both primes and targets) was coded by the first author for construction (DO, PO, or non-dative). Some participants produced fewer than 20% datives in the target ($n = 3$), or fewer than 20% of one alternant ($n = 11$), so they were excluded for producing insufficient variation. This left 21 subjects for the analysis. Prime-target pairs where the prime and the target were not both completed with a dative were also excluded ($n = 68$), leaving 100 tokens for the analysis.

The data were analyzed with mixed-model logistic regression (for more details on analyzing categorical data with such models, see Jaeger 2008). The dependent variable was whether the prime construction was repeated in the target (1 = repeated, 0 = not repeated). A positive and significant model intercept would indicate priming: it would show that prime construction affected the target construction. The independent variable was a categorical variable representing the idiomaticity of the prime. We ran a mixed-effect model with idiomaticity as a fixed effect. The model also included a random effect of subject and another random effect of item that modeled whether the primes had the same subject and recipient (16 levels, these were sometimes repeated in order to produce more natural stimuli). There was a significant effect of priming ($B = 0.63$, $p < .005$), indicating that the alternant produced in the prime was likely to be repeated in the target. There was no main effect of idiomaticity ($B = -0.03$, $p > .8$).

We also tested an effect of a continuous covariate of idiomaticity derived from the norming data because covariates have been argued to have increased power over arbitrarily defined categorical factors (Baayen 2008, p. 237). There was still no effect of idiomaticity ($B = 0.03$, $p > .8$). We also tested for main effects or interactions with alternation bias (again derived from the norming data), because Konopka and Bock found an effect of construction flexibility on priming. We found no effects or interactions with construction flexibility; however, our materials were designed to have less variability along this dimension, which may explain the difference between our results and Konopka and Bock's.

The general priming effect is illustrated in Figure 2. The y-axis shows the proportion of primes repeated for each condition on the x-axis. For ease of visualization, the idiomaticity factor is shown as a categorical variable. The effect of priming is clear in that all conditions have a repetition rate of greater than 50%: the construction is more likely to be repeated than not. Another way to quantify the effect is to see if the proportion of one construction (say PO) is higher after primes of the same construction. PO constructions were produced more often after PO primes (53%) than after DO primes (26%), whether the prime was idiomatic or not (62% PO after an idiomatic PO prime, and 43% PO after a non-idiomatic PO prime; this difference is not statistically significant).

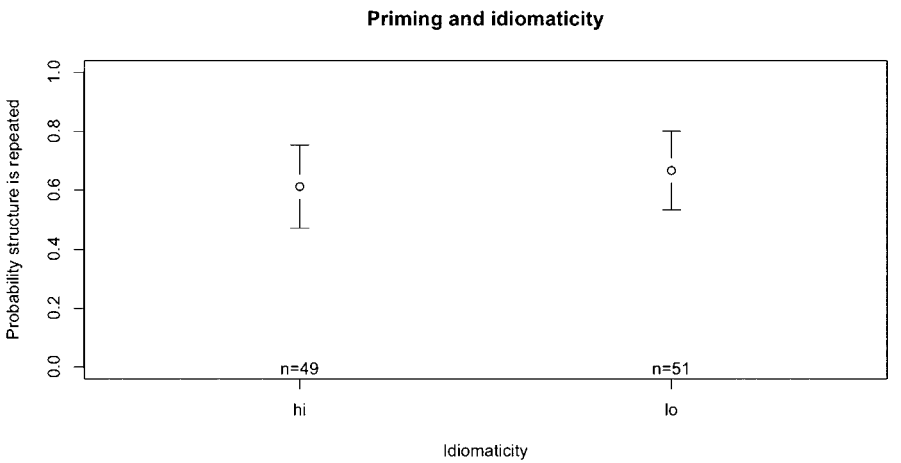


Figure 2. Proportion of prime structures repeated in the target for idiomatic and non-idiomatic primes. The bars represent 95% confidence intervals

The results of the experiment show that idiomaticity (as defined and manipulated in this experiment) does not affect priming. These findings are consistent with Konopka and Bock's finding that flexibility is the determinant of priming behavior, not compositionality.

3.7. Discussion

These results, along with those found earlier, suggest that idioms do have internal syntactic structure, and in some respects, involve syntactic processes similar to those used with non-idiomatic expressions. Semantic compositionality does not seem to determine whether an utterance is, or is not, an instance of a construction. Given that structures that belong to the same construction are more likely to prime than those that belong to two superficially similar constructions (supposedly stored separately because of the non-compositional meaning of one of them), the priming results indicate a shared construction for idiomatic and non-idiomatic datives. Despite the fact that the meaning of *The lifeguard gave the surfer the creeps* is harder to derive from its parts than the meaning of *The mother gave the hungry baby some food*, the former is still perceived to be an instance of the Double Object/Prepositional Object construction.

One important limitation of the current results is that the idiomatic forms we used (since we were limited to verbs that could alternate) were not as non-compositional as in previous work on idiom processing. However, our idiomatic items were still judged as significantly less compositional than the non-idiomatic ones in the norming study. Our items were also limited in that they all used only one verb (*give*). This could have been a problem if our main question was about the generality of priming (which has been reported with many verb types), but since we were concerned with the similarity in priming between more and less idiomatic items, the repetition of the verb becomes less of an issue. We will return to both these issues in the General Discussion.

The priming results are more consistent with a single-system model in which compositional and non-compositional phrases are processed in a qualitatively similar way, and where idioms have internal structure. This is in fact quite similar to Konopka and Bock's 'structural' model of idioms where "their internal structure is accessible to and undergoes the type of generalized syntactic processing involved in both production and comprehension" (pp. 4). There is another argument to make against using semantic compositionality as a determinant of lexical storage, as in dual-system models (Goldberg 2006; Wray 2002). As we noted earlier, it is not

easy to determine whether a phrase is compositional or not, since compositionality is more a matter of degree than a binary distinction. Moreover, from the perspective of the child learner who has yet to home in on the regularities of the language, all linguistic input starts out being idiosyncratic and 'irregular' to some degree. Starting out, a child cannot know that *dogs* is regular but *teeth* is not. To extract patterns of regularity, the child first has to have access to multiple stored tokens, both regular and irregular.

This is not to say that idiomatic expressions are not special – speakers have to acquire knowledge about their idiosyncratic semantic (and sometimes syntactic) features to be able to properly produce and comprehend them. But such knowledge may be learned via the same mechanisms used to derive meaning and structure from compositional forms. Put differently, the unique meanings (and norms of use) for idiomatic expressions can be learned from experience without blocking out what they have in common with other, more compositional phrases.

Idiomatic and non-idiomatic datives in the same construction share structural features like argument order and the presence or absence of the preposition '*to*', so they prime that structure to the same degree. Similarity is also the primary factor that drives generalization in single-system models like connectionist and exemplar models, and the finding that idioms can indeed prime argues for a model with one representational mechanism rather than a model with two separate mechanisms for compositional and non-compositional forms.

4. General Discussion

Dual-system models often use two criteria to differentiate between those structures that are stored and those that are computed: compositionality and frequency. We have presented two experiments that show that linguistic structures are processed in qualitatively the same way regardless of where they fall on the frequency and compositionality continua. In Arnon and Snider (2010), we showed that compositional 4-word phrases are responded to more quickly the more frequent they are. This is evidence that language users have knowledge about the frequency of phrases this size, just as they have knowledge of the frequency of words (regular and irregular, Alegre and Gordon 1999; Baayen et al. 1997; Taft 1979). In this respect, 'stored' elements seem no different from 'computed' ones. Importantly, we also showed that they are sensitive to frequency across the

continuum: there is no threshold beyond which phrases are attended to. High and low frequency phrases are processed in a qualitatively similar way: their processing is affected by a continuous measure of frequency.

We also presented a priming experiment that showed that compositionality is unable to differentiate stored and computed representations. Both non-compositional and compositional dative structures prime their construction, and do so to the same degree. One limitation of our results is that the idiomatic phrases in our experiment were somewhat compositional: they were not judged at the far end of the scale in the norming experiment (though they were still judged as significantly less compositional than the non-idiomatic phrases), and they were not as strongly idiomatic as in previous experiments (Konopka & Bock, 2008; Sprenger et al. 2006). While it is possible that more highly idiomatic phrases would prime less, thereby showing their diminished internal structure, such a result has not been found to date. Even studies using more idiomatic phrases than ours (Konopka & Bock, 2008; Sprenger et al. 2006) still find evidence for internal structure in idioms and strong parallels with compositional phrases. It is possible that "stored" and "computed" forms can be empirically distinguished on the basis of other measures, in particular, flexibility (which was not manipulated in the current study). But given the multitude of components that make up flexibility (morphological, syntactic and lexical, Wulff, 2008), and given the fact that it seems to have an effect on priming regardless of idiomaticity (Konopka & Bock, 2008), it is unlikely to provide a clear empirical criteria for distinguishing "stored" and "computed" forms. In future work, we would like to investigate further parallels in the processing of more and less flexible forms as well.

It is also possible that this experiment primed verb specific representations because the same verb was used in the prime and target (Gries &Wulff, 2005). However, even assuming that we are activating subtypes of the *give* dative constructions, we still manipulate compositionality within this set of constructions, so our results are not confounded by repeating *give* in prime and target. Idiomatic and non-idiomatic phrases primed (this construction) to the same degree.

The priming experiment results affirm the finding of Konopka and Bock (2009): compositionality does not affect how much that structure persists and is re-used in later processing. The priming result is also consistent with a 'weak' dual-system model (like that proposed by Konopka and Bock 2009), where there is a distinction between idioms and purely compositional phrases, but idioms are formed using 'regular' syntactic processes. However, such an account leaves little of the original concep-

tion of idioms as holistic lexical entries. Taken together with the frequency results, our findings are more consistent with the redundancy in storage predicted by single-system models.

These results support one of the fundamental tenets of single-system models: the similarity between 'stored' and 'computed' forms. In such models, similarity between structures and the frequency of those structures determines the extent to which they generalize. This is unlike dual-system models, which have two separate mechanisms, storage and computation, drawing on different representational bases. The criteria that have been argued to distinguish these two types of structures, semantic compositionality and frequency, are challenged by the current findings Neither serves as a clear empirical criterion distinguishing 'stored' from 'computed' forms. Non-compositional forms still appeared to have internal structure, and there was no evidence for a threshold beyond which frequency affected processing: more frequent structures were processed more easily across the continuum. These findings echo those in the morphological literature showing parallels in the processing of regular and irregular forms (Alegre and Gordon 1999; Baayen et al. 1997; Baayen 2006; Taft 1979).

The difficulty in finding a clear criterion for inclusion in the lexicon has led Elman (2009) to the radical solution of "lexical knowledge without a lexicon". Elman reviews numerous studies detailing the rich information language users have about verbs (from the agents they appear with to the discourse situations they evoke), and the rapid way this information is used in online processing. To explain the ready availability of such detailed, situation-specific lexical information in online processing, Elman suggests that "either the lexicon must be expanded to include factors that do not plausibly seem to belong there; or else virtually all information about word meaning is removed, leaving the lexicon impoverished". He argues for a third alternative, an emergentist model in which linguistic knowledge is viewed as a constantly changing dynamic system and where the lexicon doesn't contain fixed units but dynamic patterns. We propose that phrasal frequency effects and idiom priming effects similarly require a model that transcends traditional notions of the lexicon.

One possibility, in line with exemplar models of language (Bod 1998; Goldinger 1996; Johnson 1997; Pierrehumbert 2001, 2006) is to implement the representations produced by the exemplar-based syntactic models of Bod (1998, 2006) in a spreading-activation network, as proposed in Snider (2008). In the model that Bod presents, syntactic productivity is achieved by starting with arbitrarily large linguistic units and deducing syntactic structure from similarity and statistical inference. The resulting lexicon has

structurally analyzed chunks of different grain-sizes, which are necessarily redundant, along with a mechanism for generating larger structures out of them. The processing of units is influenced by the probability of the smaller units used to form them (Bod 2006). Implementing these representations in a spreading-activation network (Snider 2008) will result in patterns of varying levels of abstraction (from fully realized strings of words, to fully abstract constructions) that are linked to each other, and whose activation is related, among other factors, to frequency of occurrence.

Multi-word phrases can be represented naturally in this model, and be linked to the words and smaller strings they consist of. For example, the phrase *don't have to worry* would be linked to *don't*, *have*, *to*, and *worry* as well as *don't have*, *to worry*, and so on. Multi-word phrases, including idioms, are also linked to the more abstract units they are instances of: verb-phrases, constructions, etc. (so *give the old sailor a lift* is linked to the DO construction as well as its own idiom). The same would apply to all phrases, regardless of their semantic compositionality or frequency, and would lead to complementary representations at different grain sizes.

Adopting a single-system model of linguistic representation has many additional implications for language processing and learning. In comprehension, processing should take advantage of such knowledge of the likelihood of generalizations at many levels of abstraction and semantic compositionality. There is already evidence that processing is affected by expectations at many levels: the frequency of words in specific syntactic structures (verb-subcategorization biases, Clifton, Frazier, and Connine 1984; Garnsey, Pearlmutter, Myers, and Lotocky 1997; MacDonald, Pearlmutter, and Seidenberg 1994), co-occurrence relations between verbs and specific arguments (Trueswell, Tanenhaus, and Garnsey 1994); as well as the overall frequency of syntactic structure (e.g. main clause vs. reduced relative, Frazier and Fodor 1978). Representing the connections between similar structures at differing levels of semantic compositionality may play a role in the processing of metaphorical language and conventional expressions that are essential for fluent communication (Pawley and Syder 1983). Production models would have to take into account the possibility of selecting whole phrases from storage, rather than from the two levels (lexicon and grammar) of current models (and there is growing evidence that production is sensitive to fine-grained expectations, Jaeger, in press; Jurafsky et al. 2001; Gahl and Garnsey 2004; Tily et al. 2009) Phrasal storage also has implications for learning, especially if representational knowledge arises by generalizing over tokens of stored experience. Using larger units may aid in extracting grammatical regularities (e.g., using

frequent frames to learn about grammatical categories, Mintz 2003), and not doing so may be one of the factors that hinders adult language learning (Arnon and Ramscar 2009, 2012).

5. Conclusion

In this chapter we have presented findings that challenge the distinction between 'stored' and 'computed' forms by (1) undermining the empirical criteria used to distinguish between them, and (2) demonstrating parallels in the processing of words and phrases (frequency effects), and idiomatic and non-idiomatic phrases (priming). Frequency, while often thought to be a marker of lexical storage, affects the processing of compositional phrases. Idioms, while often thought to be holistically stored, show priming of their construction just like non-idioms. Together, these findings highlight the utility of models that deal with all linguistic experience in a qualitatively similar fashion, and allow for experience to influence the learning, representation and processing of all linguistic patterns.

References

Alegre, M. and Gordon, P.
1999 Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40: 41–61. Elsevier.

Arnon, I. and Ramscar, M.
2009 Order-of-acquisition affects what gets learned. In N.A. Taatgen and H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (pp. 2112–2117). Cognitive Science Society.

Arnon, I. and M. Ramscar
2012 Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition* 122: 292–305.

Baayen, R.H., T. Dijkstra and R. Schreuder
1997 Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language* 37: 94–117.

Baayen, H.
2008 *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge University Press.

Baayen, R., Feldman, L. and Schreuder, R.
2006 Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language* 55: 290–313.
Bannard, C. and Matthews, D.
2008 Stored word sequences in language learning: the effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19: 241–248.
Bard, Ellen, Dan Robertson & Antonella Sorace
1996 Magnitude estimation of linguistic acceptability. *Language* 72.1: 32–68.
Barlow, G.M. and Kemmer, S.E.
2000 Usage-Based Models of Language. Stanford, CSLI.
Bates, E. and MacWhinney, B.
1989 Functionalism and the competition model. In *The crosslinguistic study of sentence processing* 3–73.
Bell, A., Brenier, J., Gregory, M., Girand, C. and Jurafsky, D.
2009 Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60: 92–111.
Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory and M., Gildea, D.
2003 Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America* 113: 1001–1024.
Bock, J.K.
1986 Syntactic persistence in language production. *Cognitive Psychology* 18: 355–387.
Bod, R., J. Hay and S. Jannedy
2003 *Probabilistic linguistics*. Cambridge, MA: MIT Press.
Bod, R.
1998 *Beyond Grammar: An Experience-based Theory of Language*. Center for the Study of Language and Information, CA.
Bod, R.
2006 An all-subtrees approach to unsupervised parsing. Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL, 865–872.
Branigan, H.P., M.J. Pickering, S.P. Liversedge, A.J. Stewart and T.P. Urbach
1995 Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research* 24: 489–506.
Bresnan, J., Cueni, A., Nikitina, T. and Baayen, H.
2007 Predicting the dative alternation. *Proceedings of the Royal Netherlands Academy of Science Workshop on Foundations of Interpretation.* Amsterdam.
Bybee, J. and P. Hopper
2001 *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Bybee, J.
1998 *The emergent lexicon*. In Chicago Linguistic Society, 34: 421–435.
Bybee, J.
2006 From usage to grammar: The minds response to repetition. *Language* 82: 711–733.
Bybee, J. and McClelland, J.
2005 Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review* 22: 381–410.
Bybee, J. and Scheibman, J.
1999 The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics* 37: 575–596.
Cacciari, C. and Tabossi, P.
1988 The comprehension of idioms. *Journal of Memory and Language* 27: 668–683.
Chang, F., Dell, G.S. and Bock, J.K.
2006 Becoming syntactic. *Psychological Review* 113: 234–272.
Chomsky, N.
1957 *Syntactic Structures*. The Hague: Mouton.
Chomsky, N.
1965 *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.
Chomsky, N.
1981 *Lectures on Government and Binding*. Dordrecht: Foris Publications.
Chomsky, N.
1995 *The Minimalist Program*. Cambridge, MA: The MIT Press.
Cieri, C., Miller, D. and Walker, K.
2004 The Fisher Corpus: a Resource for the Next Generations of Speech-to-Text. *In Proceedings of LREC 2004*.
Clifton, C., Frazier, L. and Connine, C.
1984 Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior* 23: 696–708.
Cutting, J.C. and Bock, K.
1997 That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition* 25: 57–71.
Elman, J.L.
1991 Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning* 7: 195–225.
Elman, J.L.
2009 On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science* 33: 547–582.
Estival, D.
1985 Syntactic priming of the passive in English. *Text* 5: 7–21.
Fodor, J.
1983 *The modularity of mind*. Cambridge, MA: MIT Press.

Frazier, L. and Fodor, J.D.
1978 The sausage machine: A new two-stage parsing model. *Cognition* 66: 291–325.
Gahl, S. and S.M. Garnsey
2004 Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language* 80: 748–775.
Garnsey, S., Pearlmutter, N., Myers, E. and Lotocky, M.
1997 The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. *Journal of Memory and Language* 37: 58–93.
Gibbs, R. and N. Nayak
1989 Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology* 21: 100–138.
Godfrey, J.J., E.C. Holliman and J. McDaniel
1992 Switchboard: Telephone speech corpus for research and development. In IEEE ICASSP, 517–520.
Goldberg, A.E.
1995 *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.
Goldberg, A.
2006 *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press.
Goldinger, S.D.
1996 Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22: 1166–1183.
Gries, Stefan Th. and Stefanie Wulff
2005 Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics* 3: 182–200.
Gregory, M.L., Raymond, W.D., Bell, A., Fosler-Lussier, E., Hay, D., Pierrehumbert, J. and Beckman, M.
2004 Speech perception, well-formedness, and the statistics of the lexicon. *Papers in Laboratory Phonology* VI: 58–74.
Jackendoff, R.
1995 The boundaries of the lexicon, in M. Everaert, E.-J. van der Linden, A. Schenk, and R. Schreuder, (Eds.), *Idioms: Structural and Psychological Perspectives*, 133–165. Hillsdale, NJ: Erlbaum.
Jackendoff, R.
1997 *Architecture of the language faculty*. Cambridge, MA: MIT Press.
Jackendoff, R.
2002 *Foundations of language*. Oxford University Press, New York.
Jaeger, T.F.
2008 Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4): 434–446.

Jaeger, T.F.
in press Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*.
Johnson, K.
1997 Speech perception without speaker normalization. In K. Johnson and Mullenix (Eds.) *Talker Variability in Speech Processing*, 145–166. San Diego: Academic Press.
Jurafsky, D., Bell, A., Gregory, M. and Raymond, W.
2001 Probabilistic relations between words: Evidence from reduction in lexical production. In: J. Bybee and P. Hopper, P. (Eds.), *Frequency and the Emergence of Linguistic Structure*, 229–254. Amsterdam: John Benjamins.
Kaplan, R.M. and Bresnan, J.
1982 Lexical-Functional grammar. In J. Bresnan (Ed.), *The mental representation of grammatical relation*, 173–281. Cambridge, MA: MIT Press.
Konopka, A.E. and Bock, J.K.
2009 Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology* 58: 68–101.
Langacker, R.
1987 *Foundations of Cognitive Grammar*. Stanford University Press.
Langacker, R.
1988 A usage-based model. *Topics in Cognitive Linguistics* 50: 127–163.
Levy, R. and Jaeger, T.F.
2007 Speakers optimize information density through syntactic reduction. In: B. Schlokopf, J. Platt, and T. Hoffman (Eds.), *Advances in neural information processing systems* (NIPS), 19, 849–856. Cambridge, MA: MIT Press.
MacDonald, M.C., Pearlmutter, N.J. and Seidenberg, M.S.
1994 The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101: 676–703.
Marchman, V.A.
1993 Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience* 5: 215–234.
McDonald, S.A., and Shillcock, R.C.
2003 Eye-movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science* 14: 648–652.
Mintz, T.H.
2003 Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90: 91–117.
Monsell, S.
1991 The nature and locus of word frequency effects in reading. In D. Besner and G.W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition*, 148–197. Hillsdale, NJ: Erlbaum.

Morton, J.
1969 The interaction of information in word recognition. *Psychological Review* 76: 165–78.
Newmeyer, F.
2003 Grammar is Grammar and Usage is Usage. *Language* 79: 682–707.
Nunberg, G., Wasow, T. and Sag, I.A.
1994 Idioms. *Language* 70: 491–538.
Pawley, A. and Syder, F.H.
1980 Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J.C. Richards & R. Schmidt (Eds.), *Communicative competence*, 191–225. London: Longmans.
Pickering, M.J. and H.P. Branigan
1998 The Representation of Verbs: Evidence from Syntactic Priming in Language Production. *Journal of Memory and Language* 39: 633–651.
Pierrehumbert, J.
2001 Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee and P. Hopper (eds.) *Frequency Effects and the Emergence of Lexical Structure*, 137–157. Amsterdam: John Benjamins.
Pierrehumbert, J.
2006 The next toolkit. *Journal of Phonetics* 34: 516–530.
Pinker, S. and Ullman, M.T.
2002 The past and future of the past tense. *Trends in Cognitive Sciences* 6: 456–463.
Pinker, S.
1991 Rules of Language. *Science* 253: 530–535.
Pinker, S.
1999 *Words and Rules: The Ingredients of Language*. Basic Books.
Pinker, S. and Prince, A.
1988 On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28: 73–193.
Plunkett, K. and Marchman, V.
1991 U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition* 38: 43–102.
Plunkett, K. and Marchman, V.
1993 From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition* 48: 21–69.
Pollard, C.J. and Sag, I.A.
1994 *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Poplack, S.
1980 The notion of the plural in Puerto Rican Spanish: Competing constraints on/s/deletion. *Locating language in time and space*, 55–86.
Prince, A. and Pinker, S.
1988 Rules and connections in human language. *Trends in Neurosciences* 11: 195–202.
Reali, F. and Christiansen, M.
2007 Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language* 57: 1–23.
Roland, D., Dick, F. and Elman, J.L.
2007 Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57: 348–379.
Rumelhart, D. and McClelland, J.
1986 On learning the past tenses of English verbs. In: D. Rumelhart and J. McClelland, J. (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 216–271. Cambridge, MA: MIT Press.
Sankoff, D. and S. Laberge
1978 Statistical dependence among successive occurrences of a variable in discourse. *Linguistic Variation: Methods and Models*, 119–126.
Seidenberg, M.S.
1994 Language and connectionism: The developing interface. *Cognition* 50: 385–401.
Snider, N.
2008 *An exemplar model of syntactic priming*. Unpublished dissertation, Stanford University, Stanford, CA.
Sprenger, S., Levelt W.J.M. and Kempen, G.
2006 Lexical access during the production of idiomatic phrases. *Journal of Memory and Language* 54: 161–184.
Taft, M.
1979 Recognition of affixed words and the word-frequency effect. *Memory and Cognition* 7: 263–72.
Tily, H., Gahl, S., Arnon, I., Kothari, A., Snider, N. and Bresnan, J.
2009 Pronunciation reflects syntactic probabilities: Evidence from spontaneous speech. *Language and Cognition* 1:2.
Tomasello, M.
2003 *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
Trueswell, J.C., Tanenhaus, M.K. and Garnsey, S.M.
1994 Semantic Influences on Parsing: The use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33: 285–318.
Ullman, M.T.
2001 A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience* 2: 717–726.

Ullman, M.T., Corkin, S., Coppola, M., Hickok, M., Growdon, J.H., Koroshetz, W.J. and Pinker, S.
1997 A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience* 9: 266–276.

Ullman, M.T., Pancheva, R., Love, T., Yee, E., Swinney, D. and Hickok, G.
2005 Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language* 93: 185–238.

Ullman, M.T. and Wallenski, M.
2005 Reply: Moving past the past tense. *Brain and Language* 93: 248–252.

Underwood, G., Schmitt, N. and Galpin, A.
2004 The eyes have it: An eye-movement study into the processing of formulaic sequences. In: N. Schmitt (Ed.) *Formulaic Sequences: Acquisition, Processing, and Use*, 153–172. Amsterdam: John Benjamins.

Weiner, E.J. and William Labov
1983 Constraints on the agentless passive. *Journal of Linguistics* 19: 29–58.

Wray, A.
2002 *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wulff, Stefanie
2008 *Rethinking Idiomaticity: A Usage-based Approach*. London/New York: Continuum.

Appendix A: experimental idioms for study 2

The columns contain the following information: item number, condition, prime structure, mean idiomaticity rating, prime stimulus, target stimulus.

Item	Condition	Prime	Idiom-aticity	Prime stimulus	Target stimulus
1	nonidiom	do	2	The racing driver gave the helpful mechanic (a job)	The patient gave
	nonidiom	po		The racing driver gave a job to (helpful mechanic)	The patient gave
2	idiom	do	3.7	The efficient secretary gave the grumpy businessman (a look)	The little girl gave
	idiom	po		The efficient secretary gave a look to (grumpy businessman)	The little girl gave
3	nonidiom	do	2	The famous journalist gave the fashion designer (her address)	The diver gave
	nonidiom	po		The famous journalist gave her address to (fashion designer)	The diver gave
4	idiom	do	3.7	The blackmailer gave the sleazy journalist (control)	The lonely sailor gave
	idiom	po		The blackmailer gave control to (sleazy journalist)	The lonely sailor gave
5	nonidiom	do	2.7	The millionaire gave the struggling artist (some advice)	The explorer gave
	nonidiom	po		The millionaire gave some advice to (struggling artist)	The explorer gave
6	idiom	do	4.6	The mother gave the hungry baby (a boost)	The flight attendant gave
	idiom	po		The mother gave a boost to (hungry baby)	The flight attendant gave
7	nonidiom	do	2.5	The researcher gave the experienced surgeon (some information)	The man gave
	nonidiom	po		The researcher gave some information to (experienced surgeon)	The man gave
8	idiom	do	4.4	The cheerful engineer gave the architect (an edge)	The teacher gave
	idiom	po		The cheerful engineer gave an edge to (architect)	The teacher gave

9	nonidiom	do	1.1	The mother gave the hungry baby (some food)	The flight attendant gave
	nonidiom	po		The mother gave some food to (hungry baby)	The flight attendant gave
10	idiom	do	4.5	The famous journalist gave the fashion designer (a hand)	The diver gave
	idiom	po		The famous journalist gave a hand to (fashion designer)	The diver gave
11	nonidiom	do	2.2	The lifeguard gave the surfer (a list)	The inventor gave
	nonidiom	po		The lifeguard gave a list to (surfer)	The inventor gave
12	idiom	do	4.3	The spy gave the double agent (trouble)	The consultant gave
	idiom	po		The spy gave trouble to (double agent)	The consultant gave
13	nonidiom	do	2.4	The grandmother gave the little girl (some money)	The tennis fan gave
	nonidiom	po		The grandmother gave some money to (little girl)	The tennis fan gave
14	idiom	do	3.8	The woman gave the new neighbor (credit)	The librarian gave
	idiom	po		The woman gave credit to (new neighbor)	The librarian gave
15	nonidiom	do	2.3	The kind teacher gave the youngster (directions)	The private detective gave
	nonidiom	po		The kind teacher gave directions to (youngster)	The private detective gave
16	idiom	do	4.2	The lifeguard gave the surfer (five)	The inventor gave
	idiom	po		The lifeguard gave five to (surfer)	The inventor gave
17	nonidiom	do	2.6	The wedding planner gave the guests (a picture)	The pharmacist gave
	nonidiom	po		The wedding planner gave a picture to (guests)	The pharmacist gave
18	idiom	do	4.1	The car salesman gave the couple (some thought)	The park ranger gave
	idiom	po		The car salesman gave some thought to (couple)	The park ranger gave
19	nonidiom	do	2.6	The manager gave the secretary (instruction)	The boyfriend gave
	nonidiom	po		The manager gave instruction to (secretary)	The boyfriend gave

20	idiom	do	4.8	The lifeguard gave the surfer (the creeps)	The inventor gave
	idiom	po		The lifeguard gave the creeps to (surfer)	The inventor gave
21	nonidiom	do	2	The car salesman gave the couple (a job)	The park ranger gave
	nonidiom	po		The car salesman gave a job to (couple)	The park ranger gave
22	idiom	do	3.7	The manager gave the secretary (a shot)	The boyfriend gave
	idiom	po		The manager gave a shot to (secretary)	The boyfriend gave
23	nonidiom	do	2.1	The efficient secretary gave the grumpy businessman (an answer)	The little girl gave
	nonidiom	po		The efficient secretary gave an answer to (grumpy businessman)	The little girl gave
24	idiom	do	4.6	The captain gave the old sailor (a lift)	The bus driver gave
	idiom	po		The captain gave a lift to (old sailor)	The bus driver gave

Measuring Mental Entrenchment of Phrases with Perceptual Identification, Familiarity Ratings, and Corpus Frequency Statistics

Catherine Caldwell-Harris, Jonathan Berant and Shimon Edelman

Abstract

Novel evidence for the pervasiveness of frequency effects is provided in two studies. Perceptual identification probabilities were obtained for word pairs varying from very high frequency (*thank you*) to pairs that were merely legal (*some cans*) or random (*victim cheese*). Frequency effects were strong for recognizable collocations, but also occurred for merely legal pairs and random pairs, which had 0 or only a few occurrences in the Corpus of Contemporary American English (COCA). These results are contrary to models proposing that statistics are maintained or exemplars stored only for sequences with some minimum frequency. Perceptual identification data was compared to both Google frequencies and those from COCA and differences are discussed. Usage-based and emergentist models were also supported by findings of greater processing fluency in identifying multi-word utterances from individuals who varied in their history of exposure to these utterances. Religious Jews had better identification of phrases from daily prayers than for weekly or annual prayers. Compared to religious Jews, secular Jews had overall poorer identification of the religious phrases and showed only weak frequency effects. Frequency effects that vary according to individual differences in language exposure are a natural outcome of usage-based theories of language and are thus a promising arena for testing specific predictions about how usage influences entrenchment.

1. Introduction

Word recognition is the Petri dish of the cognitive sciences. The processes hypothesized to govern naming, identifying and evaluating words have shaped this field since its origin in the 1970s. Techniques to measure lexical processing are not just the back-bone of the typical experimental psychology

laboratory, but are now routinely used by cognitive neuroscientists to study brain processing and increasingly by social and clinical psychologists (Eder, Hommel, and De Houwer 2007). Models developed to explain lexical processing have also aspired to be statements about the nature of human cognition (e.g., connectionist models, McClelland and Patterson 2002). Words were convenient objects to study for cognitive psychologists because they are well-defined and their nature as alphabetic strings fit computational strengths of the computer programming languages of the 1970s and 1980s which excelled at string manipulation.

The long-standing emphasis on words as the building blocks of language derives from a historical allegiance to parsimony in representation, as exemplified by the words-and-rules approach to language (Pinker 1999). This approach proposes that the building blocks of language are a set of basic units (words) and rules for combining them into larger structures. Like the older proposal of "large words" as the way to explain idioms (Swinney and Cutler 1979), the words-and-rules approach specifies that non-compositional expressions are stored as unanalyzed wholes in the lexicon, while all compositional expressions, even highly frequent ones, are produced by combining words.

But are words actually the privileged unit of mental representation and processing that all of this scientific attention makes them out to be? Like a growing number of other language researchers, our answer is no (see, e.g., Bybee and Hopper 2001; Wray 2002). We propose that the mental representations form a continuum, from word combinations which have fossilized into single units (nightclub) to those that both exist as independent units and yet have bonds, varying in tightness, with the words with which they frequently co-occur (Bybee and Scheibman 1999; Langacker, 1987; Harris 1998). Multiword expressions of varying length can also become entrenched and obtain status as units (Wray 2002).

The first line of support for this view is the simple observation that fluent speakers easily recognize the familiarity and cohesive quality of word combinations in their language. Examples in English include common noun compounds (last year, brand new), verb phrases (cut down, get a hold of, faced with) and other multi-word expressions such as common sayings and references to cultural concepts (saved by the bell, speed of light; Jackendoff 1995). These frequently co-occurring word sequences, often referred to as collocations, have only recently been studied by psycholinguists. Several publications now argue that language processing and human cognition need to be expanded (if not altered) to accommodate speakers' wide-ranging

knowledge of common word combinations (e.g., Arnon and Snider 2010; Ellis 2002; Bannard and Matthews 2008).

This represents a major shift in how multiword expressions have been regarded by language researchers. Historically, idioms were regarded as a problematic phenomenon for models of the mental lexicon. But treating idioms as unanalyzed wholes or "large words" conflicted with linguistic and empirical findings from the 1980s, showing that the idioms are usually partially compositional and are often grammatically productive (Gibbs and Nayak 1989). In the last decade, multi-word-expressions have emerged as objects of inquiry in their own right, and not just as a problem for theories of lexical structure. An intriguing aspect of multiword expressions is that language users are sensitive to the frequency of common word combinations, just as they are to the frequency of single words, single letters, and letter combinations (Ellis 2002).

Frequency effects in multiword utterances

The prevalence of frequency effects in word combinations is an especially fruitful area of inquiry because it can be approached from diverse perspectives. Frequency effects are studied by corpus linguists, psycholinguists, and computational modelers. Frequency variation across utterances is of course very salient to corpus linguists. A pressing concern for these researchers has been to link this variation to other types of data, such as determining whether the most frequent use of a polysemous word or grammatical sequence is also the prototype (Gilquin and Gries 2009). Arppe et al. (2010) have argued for the need to use convergent methods, combining corpus analysis with various types of elicited data, including behavioral experiments, noting that there is "little or no understanding of how results from these different types of data inform one another" (p. 7). A second challenge identified by Arppe et al. (2010) is to establish conventions for interpreting corpus findings as cues to psychological entrenchment. This point is echoed by Gilquin and Gries' (2009) review of studies that use both corpora and experiments. While researchers typically hope for convergence, divergent outcome using different methods doesn't necessarily invalidate either finding, because specific measures are sensitive to specific linguistic activities (reading, speaking, comparing, judging etc., see Divjak 2008).

Psycholinguists are interested in both theoretical issues such as testing the "words and rules" and usage-based models, as well as applied topics concerning frequency. A recent example is how native vs. non-native speakers

vary in their histories of usage patterns. Non-native speakers may have too little experience to have built up language routines and multiword expressions and may thus rely on translating from their first language, with consequent errors and lack of native-like output (Ellis and Simpson-Vlach 2009). Computational modelers face the conceptual challenge of how to implement frequency effects while meeting other modeling goals such as inducing grammatical and lexical regularities.

How are multiword utterances mentally represented?

Two broad categories of models are available to theorists who want to address the representational questions posed by multiword expressions that vary in frequency, as discussed by Arnon and Snider (2010; see also Snider and Arnon, this volume).

The frequency threshold approach proposes that phrases of sufficient frequency have independent representation as a way of making processing more efficient. Open questions are what counts as sufficient frequency and what whether other factors (e.g., noncompositionality) play a role in establishing a multiword structure as a linked or unitized structure.

In contrast, the continuous approach proposes that every instance of usage affects processing and representation. The continuous approach is an implication of adopting an emergentist or dynamical systems framework (Ellis and Larson-Freeman 2006, 2009; Elman 1995; MacWhinney 1999). It also assumes the usage-based approach to language developed by linguists working in cognitive grammar (Langacker 1987) and functionalist grammar (Beckner and Bybee 2009; Bybee and Hopper 2001). According to the usage-based hypothesis, each use of an expression influences its entrenchment and future processing (Tomasello 2003). The difference between more and less frequent is thus one of degree, rather than specifying whether the sequence is stored vs. computed.

ADIOS, an emergentist model

Our own investigation into frequency effects in multiword utterances is broadly inspired by emergentist approaches. One specific computational model, ADIOS (Automatic DIstillation Of Structure; Solan, Horn, Ruppin, and Edelman 2005), was designed such that common word combinations are central to grammar induction. The primary goal of the research leading to ADIOS was to develop an unsupervised learning algorithm that could induce grammar from raw text. A strength of ADIOS is that the algorithm learns to represent not just the classical phrase-structure constituents of

grammars, but also the full range of multiword expressions, including those that show partial or complete productivity.

To explain how multiword expressions are useful for grammar induction, we briefly review the method that ADIOS uses to induce grammatical structures from a corpus of raw text. ADIOS can work on linguistic sequences of any size and quality, including short phrases such as those found in children's speech or adult's speech to children. The algorithm requires that text is divided into separate phrases, as is conventionally marked by sentence boundaries in text corpora or marked as separate utterances in databases of children's speech such as CHILDES (MacWhinney 1991). The algorithm begins with the first input sequence and scans the current corpus of utterances for any stored phrases that share a contiguous sequence of lexical items with this input sequence. For example, the input sequence *I saw the news today* may pull out from the corpus the utterances *the news is good* and *I read the news online these days*, all aligned on the shared subsequence *the news*. The algorithm then tests whether this shared subsequence represents a statistically significant combination, using the metric that the words occur together significantly more often than would be expected by chance. A metric for this is the surprise ratio, which measures how unexpected a sequence is, given the probabilities of its components (Barlow 1990). ADIOS used a variant of the surprise ratio, using the frequency in the local context to drive unitization. An implication is that a word sequence can have low global usage frequency yet stand out in a circumscribed set of contexts and thereby become entrenched.

To learn a grammar that can generate novel utterances rather than merely extracting regularities, the criterion that governs matching in ADIOS is relaxed to allow a local mismatch in the shared subsequence. By allowing local mismatches, the phrases *I heard the big news* and *I heard the latest news* would match. Allowing local mismatches has the result that words occupying the corresponding slots in the aligned phrases will be interchangeable. For example, *I read the news* and *I saw the news* will match. Multiword sequences can also match single words, allowing modifier + noun combinations to match a single noun. This matching process creates equivalence classes, resulting in increasingly abstract classes of words and multiword expressions, ultimately leading to noun and verb phrases that can vary in size. The procedure of scanning a corpus for recurring expressions is repeated recursively until no new collocations are found. The resulting set of equivalence classes can be understood as a grammar that represents both the sequential order of lexical items and collocations of varying lengths.

ADIOS has proven to be effective at grammar induction (see Waterfall, Sandbank, Onnis, and Edelman 2010), but still unknown is how closely algorithms like ADIOS match behavioral data on frequency effects. It may be premature at present to model behavioral data because of lack of knowledge about boundary conditions. For example, it is currently unknown whether some minimum frequency is required for unitization.

Technically, ADIOS could be regarded as an example of the frequency threshold approach, since unitization of a multiword utterance only occurs if that utterance occurs more frequently than would be expected by chance in a given context. But ADIOS nevertheless bears the primary hallmarks of emergentist models. Grammatical categories (and all linguistic structure) emerge from statistical processing over input sequences, and every occurrence of an utterance influences the calculations that determine unitization. In addition, there is no single threshold that determines when frequency starts influencing processing. Most importantly, even common word combinations that have objectively low frequency become unitized in ADIOS, including what would be considered "merely legal" combinations such as *the latest news* or even *the news*.

ADIOS suggests some testable hypotheses that are relevant to current knowledge gaps. Arnon and Snider (this volume) have provided evidence against the threshold approach by reporting frequency effects for more than two frequency groupings. For their response latency data, a continuous measure of frequency always had a superior fit than a dichotomous (low/high) measure, a finding consistent with continuous models. We want to push this a step farther by studying more of the frequency spectrum. ADIOS predicts frequency effects for low frequency collocations, and thus in Study 1, we looked for frequency effects in low frequency collocations, not just high frequency collocations. Since ADIOS predicts that fully compositional sequences are stored, we looked for frequency effects in modifier + noun patterns such as *her list*, *some cans* and *green skirt.*

ADIOS assumes that efficiency of processing a sequence is a function of quantity of exposure for individual learners. This encouraged us to investigate, in Study 2, whether speakers with different types of exposure to phrases would be differentially efficient at processing those phrases. Speakers who have more experience with some specific expressions should perform more efficiently in identifying those expressions than speakers with less exposure/use of those word sequences.

Wherever possible, we also wanted to address open questions about the processing of multiword expressions. Corpus linguists want to connect corpus frequency data to the results of behavioral experiments, and have

urged psycholinguists to use a diversity of paradigm (Arppe et al. 2010). Our two studies used a task which has not been previously used to study processing of multiword utterances, the perceptual identification task. In the next section we provide background about this task.

The perceptual identification task

In a standard perceptual identification task, a stimulus is briefly displayed on a computer screen, typically for durations of 30 ms to 100, and then masked with a visual noise pattern, which disrupts continued processing (Carr, 1986; Ratcliff and McKoon 1997). The subjective experience of respondents may be of seeing an unknown word, but with a few practice trials participants feel they can guess and are often correct or close. Exposure durations are usually long enough so that stimuli can be consciously perceived, but short enough that response accuracy is below ceiling. Brief display and masking makes the recognition task difficult. The difficulty is reduced, and recognition enhanced, if participants can easily match the brief, degraded input to a representation in long term memory. The stronger the long term memory representation, the more accurate is identification. Classic phenomena using the perceptual identification task include the word superiority effect, in which observers can more easily identify a word than a nonword (Rumelhart and McClelland 1981). A variant of the standard task is to display multiple words sequentially (as done in Caldwell-Harris and Morris 2008), but the dependent measure remains the same, which is to identify some or all items in the perceptual display.

2. Frequency effects in low frequency word pairs

The growing interest in relating behavioral measures to results of corpus analysis (e.g., Gilquin and Gries 2009), and the public availability of high quality corpora such as the Corpus of Contemporary American English (COCA; Davies 2010), opens the door for psycholinguists to draw on a rich source of data about frequency effects in multiword utterances: data from existing experiments can be re-analyzed by connecting performance data to corpora, as suggested by Gries 2010. Here we use COCA frequencies to analyze existing behavioral data from a perceptual identification task.

Caldwell-Harris and Morris (2008) identified a temporal illusion produced when observers perform perceptual identification on familiar word combinations. When the word combinations were highly frequent, but

presented sequentially in reverse order (i.e., *code* followed by *zip*), observers reported perceiving the familiar word pair *zip code.* For exposure durations ranging from 30 to 105 ms. for each word, observers spontaneously reversed word pairs such as *fees legal* and *step next*.

Report of words in familiar order persisted even when observers were informed that some words would be presented in reversed order and that it was important to report the order in which words appeared, even if this was not the familiar order. Participants claimed that their subjective impression was that a reversed pair such as *fees legal* had been displayed in its canonical order (*legal fees*). This impression held up even for experienced observers such as laboratory assistants who were familiar with the words on the stimulus list. We will refer to these as reversal errors, but of course they are errors only from the standpoint that observers are not sensitive to the order of word presentation, but are instead reporting the words in their most frequent order. As discussed further in that paper, this performance could be seen as optimal from a Bayesian perspective, since the prior probability of *card credit* as an independent two-word display is much lower than the probability of *credit card.*

Reversal errors were most commonly made for high frequency collocations (*keep track*, *fan club*), followed by low frequency collocation (*machine gun, any clues*), and were least common for adjective + noun combinations (*huge church*, *real skin*). Perceiving veridical order was highest for the random word pairs (*look fever*, *puppy hill*). The ability to correctly recognize the component words regardless of their order was strongly influenced by the frequency category of the word combination, and only minimally influenced by the frequency of the individual words in the string.

The data set included word pairs across the frequency range, from the highest word pairs in COCA (*thank you*) to those with low and zero frequency. Our goal in the current analysis is to connect recognition accuracy of the word pairs to three types of familiarity ratings, and two frequency corpora, Google and COCA, thereby responding to corpus linguists' plea for more work linking different types of frequency measures with different types of behavioral measures (Arppe et al. 2010; Gilquin and Gries 2009).

Description of the stimuli

The inclusion of random word pairs and the merely legal word pairs in the Caldwell-Harris and Morris (2008) data allow us to investigate whether frequency effects exist even for word pairs of absent and low frequency. Because the study was designed and administered from 1997–1999, collo-

cation frequency was determined at that time by a corpus consisting of electronic newsgroup postings used for the HAL project (Hyper Analog to Language, Lund and Burgess 1996; in 1997 Kevin Lund generously gave the authors a list of all word pairs that occurred more than 5 times in the HAL corpus). High-frequency word pairs had a mean frequency of 3700 in the 300 million word corpus, while low frequency pairs had a mean frequency of 20.6. Adjective + noun combinations were selected to be legal combinations but 0 frequency, and thus they did not exist in the HAL corpus, but were constructed to avoid violating semantic constraints, following Pustejovsky's (1995) description of semantic domains. The random word pairs were mainly noun-noun pairs which violated semantic domains, and could not easily be assimilated to an adjective + noun combination, or to any easily identifiable legal syntactic grouping, although considerable variation resulted. Stimuli, corpus frequencies, ratings and recognition accuracy appear in the Appendix.

Intercorrelations between familiarly and frequency

Google frequencies were obtained by placing quotation marks around each word pair. Frequencies ranged from a low of 74 for the random pair *weep job* to a high of 819,000,000 for the high frequency collocation *health care* (*thank you* was a close second in frequency; the high frequency of *health care* is probably an artifact of the heavy use of this phrase when Google frequencies were collected in July 2007). For COCA, a corpus of 410 million, *thank you* was our most frequent word pair, at 77,530 instances, and *health care* the 4th most frequent of our word pairs, at 28,620). Log frequency was used in graphs and calculations.

Subjective familiarity ratings for the 160 word pairs were obtained from 22 undergraduates. Raters used a 5-point scale extending from very unfamiliar to very familiar. Raters were additionally given the option of evaluating a phrase as "does not make sense." Phrases so rated were scored as 0, resulting in a familiarity scale ranging from 0 to 5.

The correlation between Google and COCA frequencies was high ($r = .93$). Correlations obtained separately just on the collocations and legal pairs were still high despite reduced range ($r = .80$ and $r = .84$ respectively). The correlation between Google and COCA was the lowest for the random pairs, $r = .60$, because 31 of the 40 random pairs had frequencies of 0 in COCA.

The correlations between corpus frequencies and familiarity ratings were also high (both $r = .88$; see top panel of Table 1). To better understand

Table 1. Correlations

Correlations Between the Two Corpus Frequencies, COCA and Google

	N	*r*	Statistical significance
All items	160	**0.93**	$p < .0001$
Collocations	76	**0.80**	$p < .0001$
Legal Pairs	44	**0.84**	$p < .0001$
Random Pairs	40	**0.60**	$p < .0001$

Correlations Between Familiarity Ratings and Corpus Frequencies

	N	COCA	Google	Stat. significance
All items	160	**0.87**	**0.88**	both $p < .001$
Collocations	76	**0.47**	**0.46**	both $p < .001$
Legal Pairs	44	**0.44**	**0.52**	both $p < .005$
Random Pairs	40	0.04	**0.31**	Google, $p < .05$

Correlations Between Perceptual Identification and Familiarity/Frequency

	N	COCA	Google	Fam.	Stat. significance
All items	160	**0.59**	**0.61**	**0.61**	all $p < .02$
Collocations	76	**0.22**	0.15	**0.34**	Fam, $p < .005$; COCA, $p = .057$
Legal Pairs	44	**0.32**	**0.34**	0.22	both $p < .03$
Random Pairs	40	0.23	**0.39**	0.29	Google, $p < .02$

Note: Bold *r* values are statistically significant.

how familiarity judgments relate to corpus frequencies, we graphed familiarity ratings as a function of COCA frequencies. As shown in Figure 1, a floor effect occurred for the random pairs. The dense cluster in the lower left side of graph occurred because 31 of the 40 random pairs were absent from COCA. Many random pairs had ratings near 0 because the majority of raters judged them non-sensical. The graph also shows that some of the legal pairs overlapped with the collocation frequency range, and considerable overlap existed in COCA frequency between the low and high collocations. It was thus decided, for the remaining analyses, to eliminate the low/high frequency division that had been determined using the HAL corpus, and to reclassify four collocations as merely legal combinations. These were the 4 that had the lowest frequency, and which met the criterion of being a legal constituent and lacking a strong idiomatic quality (the 4 word pairs were *mind bomb*, *small fuss*, *pay rate*, *sale ends*).

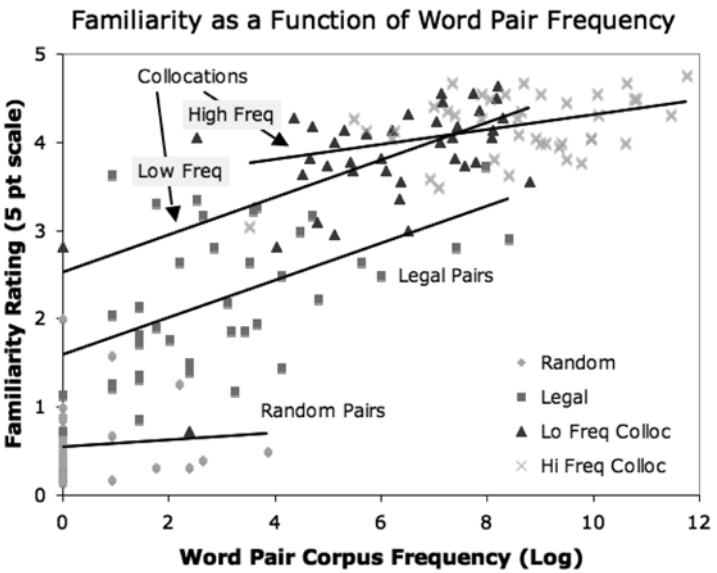


Figure 1. The relationship between students' familiarity ratings of the 160 word pairs, and the pairs' frequencies in COCA (Davis 2010), with collocations grouped according to low and high frequency. Labels and arrows are used to indicate trends lines drawn through data points, which are also labeled in the legend

The relationship between familiarity ratings and Google frequencies (not shown) is broadly similar to the relationship between familiarity and COCA frequencies, as would be expected by the overall $r = .93$ between COCA and Google frequencies. A difference is that the use of Google as the frequency metric produced a scatter plot that, compared to Figure 1, is more extended for the random pairs. The top panel of Table 1 provides correlations separately for familiarity and the two frequency metrics, for each type of word pair. Familiarity ratings for the random pairs were not related to COCA frequency ($r = -.02$), but were weakly related ($r = .31$) to Google frequency. One of the differences between COCA and Google is that all the random pairs had an existence in Google – and apparently not a random existence, ratings correlated with Google frequencies.

It is interesting that familiarity ratings for legal pairs were related to corpus frequencies, although moderately, and indeed, correlations for the legal pairs were similar in magnitude to the correlations for the collocations. The plot in Figure 1 and correlations in Table 1 thus support our

first goal, which was to show frequency effects on behavioral responses outside of high frequency correlations. Familiarity ratings correlated with corpus frequencies not just for the cases that would be predicted to do so by most theories, i.e., well-known collocations, but also for adjective + noun pairs which are generally considered compositional (e.g., *early change*, *some cans*), and randomly combined word pairs.

The next section analyzes our second behavioral measure, perceptual identification of the word pairs.

Correlations between perceptual identification and frequency/familiarity

Participants' ability to identify the words (perceptual identification or PID task) correlated overall relatively strongly with COCA ($r = .58$) and Google corpus statistics ($r = .61$). Obviously because of restricted range, correlations for the word types analyzed separately were weaker, but importantly, for the legal pairs, PID correlated with COCA ($r = .33$) and Google frequencies ($r = .34$; see bottom panel in Table 1). For random pairs, the highest correlation ($r = .39$) was with Google frequencies. Lower r values for COCA and familiarity likely resulted from floor effects. As noted above, for familiarity, floor effects occur because raters had the option of circling "doesn't make sense." Future work could explore how sensitive raters are to variations in meaning and familiarity within the category of words which are ostensibly unrelated but which can occur in print adjacent to each other, as demonstrated by above 0 Google frequencies. The finding that probability of perceptual identification for these low frequency and 0-frequency items is related to Google frequencies (and more weakly, to COCA frequencies, see Figure 2) is thus particularly impressive.

To graphically depict these frequency effects, Figure 2 plots perceptual identification as a function of COCA log frequency. For purposes of illustration, word pairs within each category were split into low and high frequency categories using a median split. The mean of the log frequencies for both low and high are plotted with standard error shown with error bars. This shows that there was considerable overlap at the high end of the random and at low end of the legal pairs in both COCA frequencies and perceptual identification.

Many of the random pairs are so nonsensical that in Google they mainly occur separated by punctuation or graphical white space (e.g., *weep job* straddles a period and references the Biblical parable of Job). Some can be assimilated to an adjective noun construction, e.g., *belt trade, trick boy*, and these had higher Google frequencies and better perceptual

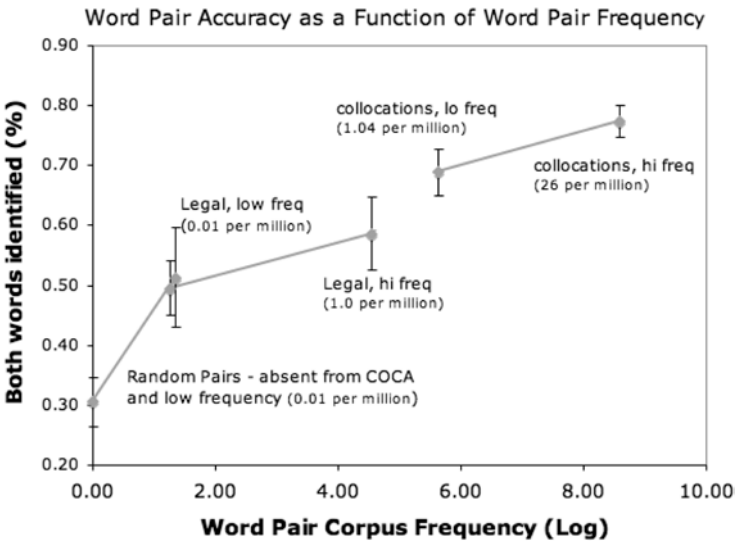


Figure 2. Perceptual identification plotted as a function of COCA frequencies, with low and high frequency categories defined by COA frequencies

identification. Future work will need to determine whether frequency effects for random pairs reduce to a difference between two types of random pairs: those that can and can't be readily assimilated to a grammatical pattern.

3. Frequent recitation creates 'tracks in the mind'

Above we tested the usage-based hypothesis and emergentist accounts by demonstrating frequency effects for word pairs across the frequency continuum, from high frequency collocations, to low frequency collocations, to merely legal combinations, to random word pairs. Another prediction of the usage-based model is that language users who have more experience with specific linguistic stimuli will have more efficient processing of those stimuli. Psycholinguists have not measured individual differences in stimulus expertise as routinely as have cognitive neuroscientists who have shown, for example, how expertise with specific objects influences brain organization (Bukach, Gauthier and Tarr 2006). One challenge is identifying speakers who reliably differ in their language experience. Language researchers have examined variation in exposure to language by comparing native vs.

second language learners (e.g., Ellis and Simpson-Vlach 2009; Gilquin and Gries 2009). Studying second language learners is certainly a good way to identify groups with more vs. less usage, but many aspects of language use are altered for non-native speakers in addition to reduced usage. It would thus be ideal to find groups who reliably vary primarily in their usage of specific expressions. An example of prior work which did this in a compelling manner is Stadthagen-Gonzalez, Bowers, and Damian (2004), who used professional expertise to detangle frequency effects from age-of-acquisition effects, arguing that acquisition age would presumably be similar for a word like *cognition* for both chemists and psychologists, but only high frequency for the psychologists.

Frequency differences in prayer habits

Our interest in collocations and routinized patterns suggested that phrases from religious rituals would have different patterns of use across groups with different prayer habits. Observant Orthodox Jews are required to recite three prayers every day. The linguistic sequences in these daily prayers would presumably be quite entrenched, compared to weekly and annual prayers. By comparing Orthodox Jews to secular Jews (and also directly inquiring about prayer recitation practices) one would have two groups with different usage patterns.

Studying Jewish prayers is a particularly fertile area because daily, weekly and annual prayers exist. We studied phrases from weekly and annual prayers, with the proviso that using such phrases is highly exploratory. Frequency may not be the most important factor for entrenchment because some prayers may have greater emotional resonances than other prayers. Weekly prayers recited on Saturdays are longer than daily prayers and occur with a different service. The annual prayers recited over the High Holy Days are further prolonged and the services carry a higher emotional charge than do services accompanying the daily and weekly prayers.

Given that one third to one-half of the Jewish population in Israel is secular, nonreligious Israelis could readily be recruited as a comparison group. Secular people generally do not recite the daily or weekly prayers, but many do attend the annual services during the High Holy Days.

Method

Participants self-identified as religious ($N = 32$, 19 females and 13 males) or secular ($N = 19$, 11 females and 8 males). Each participant completed a questionnaire detailing praying habits (frequency of praying and whether

in private or at synagogue). We additionally measured participants' knowledge of Jewish prayer texts using a phrase completion test. This test consisted of 17 phrases taken from various Jewish prayers. For example, the phrase *barux shem k'vod* ____ would be finished with *Malchuto L'olam Va'ed* (meaning of whole phrase: *Blessed is the name of his glorious kingdom for all eternity*). Phrases that were left blank or were completely wrong received 0 points, one point was given for partial completion, and two points were given for perfect completion of the phrase.

Materials for the perceptual identification task were six types of phrases which were selected to have comparable semantic and syntactic complexity (see Table 2). Religious phrases were categorized according to frequency of recitation (daily, weekly, and annual). Nonreligious phrases were selected to be either common or rare. The common phrases were drawn from Israeli culture and included political slogans, names of famous TV shows, and popular songs. The rare phrases were selected from modern Hebrew literature and poetry. Google counts confirmed that the phrases in the rare group were substantially less common than phrases in the common group (mean log frequency of rare phrases = 1.6, common phrases = 4.6; $p < 0.0001$). The sixth group was constructed out of words that appear separately from each other in Jewish prayers and do not form cohesive phrases when mixed. Each phrase group comprised eleven 2-word phrases and four 3- word phrases. The length in characters of phrases was similar across all categories (mean = 12.1, std = 0.55).

To avoid floor and ceiling effects in the perceptual identification task, exposure durations were set individually for each participant, based on performance in practice trials, with the average exposure duration for two-word phrases 71 ms and for three-word phrases, 90 ms. In order to obtain as much information as possible about participant's perception of

Table 2. Example Stimuli for Jewish Prayers Study

Phrase Type	Example and translation	
Daily prayers	*morid hatal*	he who makes the dew drop down
Weekly prayers	*nafshi yeshovev*	will exhilarate my spirit
Annual prayers	*bnei maron*	sheep and goats (archaic)
Common phrases	*shalom xaver*	good bye, friend
Rare phrases	*divrey rahav*	words of arrogance
Random	*zore'a ha'amim*	sower of nations

the stimuli, we scored their accuracy on a 3 point scale for no words correctly reported, partial correct report of the phrase (at least one word correct), and complete report of the target phrase.

Results and Discussion

Separate ANOVAS were carried out on the religious phrases, which had three frequency levels (daily, weekly and annual, see Figure 3) and the secular phrases, which had two frequency levels (common and rare, see Figure 4). For the religious phrases, religious participants were more accurate and showed stronger frequency effects, compared to secular participants. In the ANOVA with religious group and type of phrase as between-subject and within-subject predictors, main effects were obtained in the expected directions for religious group, $F(1,47) = 15.5$, $p < 0.001$, and frequency of phrase, $F(2,96) = 33.7$, $p < 0.0001$). In the ANOVA on secular phrases,

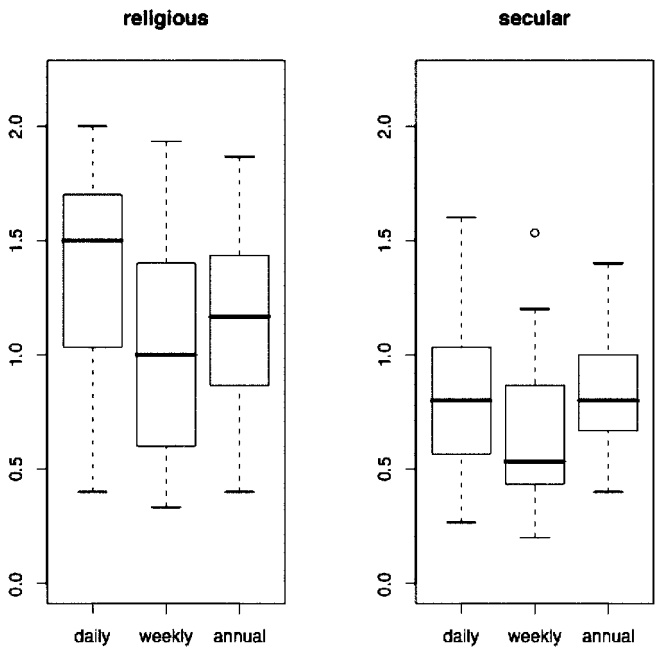


Figure 3. Mean correctness for religious participants compared to secular participants for phases taken from daily, weekly and annual prayers. Compared to secular participants, religious participants identified more words from the phrases, and showed stronger frequency effects

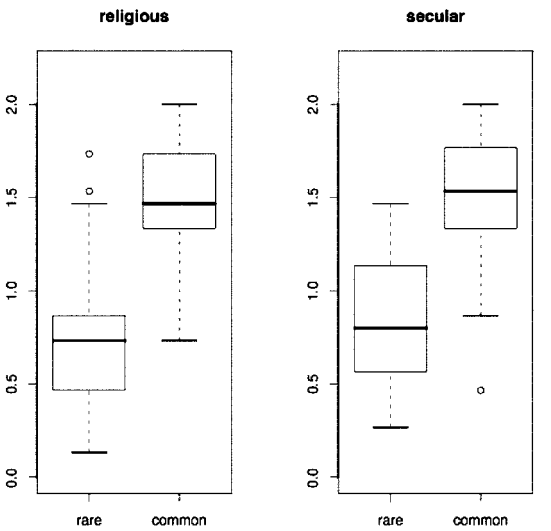


Figure 4. Correctness for nonreligious phrases did not vary according to individual religiousness; but both groups of participants more accurately identified the more frequent phrases

the two groups did not differ in their accuracy for the common and rare phrases, $F < 1$, but both groups showed frequency effects, as revealed by the strong main effect of frequency of phrase, $F(1,48) = 470.7$, $p < 0.0001$ (see Figure 4). Additional exploratory analyses, including analyses of gender effects, phrase completion scores, and self-reported prayers habits, are reported separately (Berant, Caldwell-Harris and Edelman 2008).

Phrases from daily prayers had higher accuracy than phrases selected from weekly and annual prayers, and these frequency effects were stronger in religious participants. What had not been predicted is that non-religious participants were also affected by the frequency of phrases, but in an attenuated and different manner, as indicated statistically by the group X frequency interaction, $F(2,96) = 5.5$, $p < 0.01$. For secular participants, accuracy was highest for the daily and annual prayers, and lowest for the weekly prayers (illustrated in Figure 3). It is possible that the weekly phrases were unintentionally more difficult than the daily and annual phrases. The explanation we favor is that annual phrases were more entrenched than would be expected by a once-yearly recitation, because of their high emotional charge. This can explain the relatively good accuracy for the annual

prayers shown by religious participants but is probably especially true for the secular participants. Secular participants may attend synagogue during the annual High Holy Days, an occasion that is memorable.

This study of perceptual identification of phrases from Jewish prayers directly supports the usage-based model of human language. Individuals who have greater experience with specific linguistic expressions had greater accuracy at reading the briefly displayed phrases, consistent with the predictions from ADIOS that frequently encountered sequences lay down 'tracks in the mind.'

4. General Discussion

New evidence for the pervasiveness of frequency effects

The analysis of word pair data demonstrated frequency effects not just for high frequency common word combinations, but for low frequency collocations, and for word pairs which are merely legal combinations (*some cans*). Frequency effects were found even among two word sequences that had been randomly put together and had zero frequency in COCA. This finding thus moves beyond the results of Arnon and Snider (2010) who found that response latencies varied continuously across a range of low and high frequency collocations (four-word sequences). The low frequency stimuli in Arnon and Snider's study had a minimum occurrence of 1 per million and extended to 9 per million. Our legal pairs averaged .35 per million, and our random pairs had an average frequency of .003 per million, occurring on average only 1.3 times in the 410 million word COCA. Models which propose that statistics are maintained or exemplars stored only for sequences with some minimum frequency will find it difficult to account for these frequency effects.

The goal in our study of processing of Jewish prayer phrases was to determine if frequency of uttering specific verbal expressions reliably led to processing differences, as predicted by the usage-based hypothesis and emergentist models like ADIOS (Solan et al. 2005). Religious Jews had better identification of phrases from daily prayers than for weekly or annual prayers. Compared to religious Jews, secular Jews had overall poorer identification of the religious phrases and showed only weak frequency effects. This is strong support for the usage-based hypothesis. We hope these results encourage other researchers to undertake individual differences research.

Why are language users sensitive to the frequency of word sequences?

Strong and diverse effects of frequency were found across these two studies. Why does the brain keep track of these statistics? Are frequency statistics useful for a real task in comprehension or in production, or are they a by-product of something else?

Frequency statistics are a key part of expectation-based models of comprehension (Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy 1995; Hale 2003). Humans plausibly have statistical information about the frequencies of word combinations because they store exemplars. Storing exemplars aids both acquisition and processing. As experience with the ADIOS algorithm shows (as in similar computational models, e.g., Bod, 1998), statistics about word sequences are essential to the ability to infer grammatical structure (Solan et al. 2005). Moreover, processing benefits are likely to result when people can rely on stored constructions (e.g., Lewis and Vasishth 2005), especially for highly proficient language users. Storing common phrases means that listeners can anticipate upcoming words, allowing top-down expectations to clean up a noisy speech signal or infer the completion of a sentence when only the first part has been received. More generally, storing frequency-weighted exemplars helps in predicting the world, a crucial information processing strategy that language processing shares with much of the rest of cognition (Edelman 2010).

How are word pairs stored?

One can imagine a phrasal level of representation in which top-down activation from the phrasal level explains why collocation status strongly influenced accuracy of word identification. This could be analogous to the word superiority effect, in which orthographic regularities in an individual word facilitate letter recognition (e.g., McClelland and Rumelhart 1981). Top-down and bottom-up interactive-activation and competition could explain how collocation frequency aids recognition of words which are displayed in a rapid sequence. In the case of word pairs, statistical regularities between word pairs could boost recognition of the individual words.

Algorithms like ADIOS (Solan et al. 2005) augment the interactive activation account by specifying a computational procedure whereby an initially flat representation of utterances as strings of words becomes hierarchical with experience, with collocations being assigned their own units.

Many psycholinguists propose that regularity and entrenchment in language is a matter of degree (McClelland and Patterson 2002). A continuum of unitization may exist, with fully fused word pairs like *blackboard* at the

one extreme end, *middle name* and *last chance* occupying an intermediate position and rare and novel combinations at the non-entrenched end of the continuum (Harris 1998; Wray 2002). In computational modeling, gradedness is a more revolutionary concept, and indeed many models implement unitization. The original interactive activation model (McClelland and Rumelhart 1981) assumed unitization, and ADIOS also currently assumes unitization, as noted earlier. Computational modelers need to grapple with and surmount the challenges of implementing gradedness.

Converging corpora and behavioral data

In the current paper we answered the plea of Arppe et al. (2010) and Gilquin and Gries (2009) for converging data, including relating different frequency measures to each other. The study of word pairs compared two behavioral measures (familiarity and perceptual identification) and employed two frequency corpora, COCA and Google. Very high correlations were obtained between COCA and Google. It is noteworthy that the correlation between COCA and Google was high even for sequences which are not typically considered multiword utterances, the legal pairs, ($r = .84$), and was at least moderate for random pairs ($r = .60$). This indicates that these word pairs have reliable frequency statistics in the world of printed materials. The correlations between corpus frequencies and familiarity ratings were also strong, especially when calculated across the frequency range (e.g., $r = .87$ for all 160 items). This extends to two-word pairs the observation of Balota, Pilotti and Cortese (2001) that native speakers can reliably estimate words' relative frequencies.

Why is familiarity a better fit to corpus frequencies than perceptual identification? Corpus frequencies and subjective familiarity performed similarly in predicting the perceptual data from the fast pairs paradigm described above. However, perceptual identification had smaller correlations with objective frequency than did subjective familiarity. We speculate that familiarity ratings emerge from a process that normally has sufficient time to settle into a stable state, while perceptual phenomena are influenced by more random variables, such as attention and momentary physiological factors.

It is worth noting that corpus frequencies will be a flawed estimate of mental entrenchment for words and phrases which occur very frequently in a few types of texts, causing their frequency to be as high as more general words whose occurrence is more evenly dispersed across many texts (Gries, 2010). Gries (2010) has discussed diverse aspects of this problem and has

shown that measures of dispersion improve the correlation between frequencies and response times in word recognition studies. Dispersal differences have implications for entrenchment. Psychologists have known for decades that evenly spaced learning, compared to massed learning of equal duration, results in higher test accuracy and longer retention, although several aspects of this phenomenon remain pooly understood (see Donovan and Radosevich, 1999). Future work needs to examine whether low-frequency specialized phrases have the entrenchment status that corresponds most closely to their raw frequency, or to the higher frequency of their specialized contexts.

5. Open questions and future work

Merely legal and barely legal: Why are frequency effects obtained?

Future work on processing of merely legal word pairs such as *her list* and barely legal items such as *edit center* can determine why frequency effects occur for these items. The goal of the current analysis was to determine the extent of frequency effects, and the data set was not constructed to test causality. Items in the random group such as *butter ace* and *cast bark* received below zero familiarity ratings, meaning more than half of raters judged them to make no sense. These items were absent in COCA, had low frequency in Google, and also had poor recognition on the perceptual identification task. Consider random pairs with above zero familiarity ratings such as *work use* and *edit center*. These can be easily assimilated to a modifier-noun category, and also had higher corpus frequencies and perceptual identification scores than other random pairs. Being assimilable to an adjective-noun category could cause raters to avoid using the "doesn't make sense" label. Systematic analysis of contexts across a large set of items is required to determine if these mostly appear as adjective + noun pairs, of if they are appearing in other syntactic contexts, e.g., from Google, *How to edit Center Ring*.

Does merely appearing in contiguous order influence mental entrenchment, even if the word pair is not a constituent? For example, a random pair like *city away* is not assimilable to any syntactic constituent, but gains its occurrences in corpora in sequences like *Flood of Complaints Washes Tent City Away* (from Google). Are raters reliably sensitive to the semantic features in *city* such as geographical location which match features in the adverb *away*? We hypothesize that an off-line rating may be more sensitive

to semantic feature match, while the ability to perceive a phrase under degraded conditions is more sensitive to prior experience of contiguity. This hypothesis needs to be empirically tested.

Is constituency more important than mere contiguity?

Computational models of language acquisition like ADIOS assume that language learners initially collect statistics about word and phrase co-occurrences, in order to infer grammatical constructions via their distributional regularities. What remains unclear is what types of exemplars and frequencies are retained once typical phrase boundaries are learned. Do speakers shift to partly or fully collecting statistics that respect phrase boundaries?

What is the role of semantic meaningfulness?

Corpus frequencies will be only one factor influencing entrenchment and mental representation, as has been discussed by many authors (Snider and Arnon this volume; Caldwell-Harris and Morris 2008; Gilquin and Gries 2009). Other factors may be semantic coherence, grammatical constituency, and emotional resonance. Ellis and Simpson-Vlach (2009) reported that mutual information (a statistical metric, Barlow 1990) was a better predictor of native-speakers' judgments of multiword expressions. For polysemous words, concrete senses (e.g., the "hand over" sense of *give*) are the most salient to raters, but phraseological uses (*give me a smile*) are most frequent in the corpora (Arppe et al. 2010).

It is a pervasive feature of language that frequent contextual meanings accrue to common word combinations occurring at every level of analysis (Langacker 1987). An example at the morphological level is that the common referent of *stapler* is not an object that staples but a specific instrument with recognizable shape. At the sentential level, *She felt the baby kick* typically refers to a pregnant woman feeling the kick of a fetus. As in these examples, it does seem that additional semantic coherence accrues to many commonly occurring multi-word utterances, a phenomenon discussed by Arnon and Snider (2010; see also Snider and Arnon, this volume).

When native speakers rate the familiarity of common word combinations, they may be influenced by emotional resonances of the overall meaning of the words, leading them to rate emotional phrases such as *child abuse* and *caring words* as more familiar than objectively more frequent phrases such as *rather than*. The current corpus of 160 word pairs contained collocations

such as *death bed*, *face value* and *upper hand* which contain an idiomatic quality. That is, the specific meaning of the combination is not fully predictable from component words, or at least the word pair identifies referents beyond what would be inferred on an adjective + noun analysis (e.g., *gold medal* is a specific type of award, not just a medal that is gold; *black hole* is not just a hole that is black). While semantic coherence was not a requirement for selection in the current corpus, most of the collocations in the word pair study are either idiomatic or have specific culturally acknowledged referents, while most of the legal pairs lacked this: compare the legal pair *blue wall* to the collocation *green light.* It is possible that this extra referentiality of the collocations and/or extra emotionality provides them with a recognition boost (see higher slope for collocations in Figure 2).

The data from the Jewish prayers study also contained tantalizing hints that multiple factors influence entrenchment, in addition to frequency. Secular Jews recognized annual phrases with the same accuracy as daily phrases, with weekly prayers having lower accuracy. We speculate that daily prayers benefit from being commonly known because they are supposed to be recited daily, and secular Jews may are aware of them as part of being familiar with Jewish cultural knowledge. But annual prayers benefit from the emotional charge of the High Holy Days. Future work on within-speaker variation in mental entrenchment can investigate whether "tracks in the mind" are mainly influenced by amount of exposure vs. personal emotional response to the stimuli.

Acknowledgments

We thank Mark Davies for providing us with the COCA frequencies for the 160 word pairs in the Appendix.

Bibliography

Arnon, Inbal and Neal Snider
2010 More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62: 67–82.

Arppe, Antti, Gaetanelle Gilquin, Dylan Glynn, Martin Hilpet and Arne Zeschel
2010 Cognitive corpus linguistics: Five points of debate on current theory and methodology. *Corpora* 5: 1–2.

Balota, David A., Maura Pilotti and Michael J. Cortese
2001 Subjective frequency estimates for 2,938 monosyllabic words. *Memory and Cognition* 29: 639–647.

Bannard, Colin and Danielle Matthews
2008 Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19: 241–248.
Barlow, Horace B.
1990 Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research* 30: 1561–1571.
Beckner, Clay and Joan Bybee
2009 A usage-based account of constituency and reanalysis. *Language Learning* 59: 27–46.
Berant, Jonathan B., Shimon Edelman and Catherine. L. Caldwell-Harris
2008 Tracks in the mind: Differential entrenchment of common and rare liturgical and everyday multiword phrases in religious and secular hebrew speakers. Presented at the *Annual Meeting of the Cognitive Science Society*, Washington, D.C., July 2008.
Bybee, Joan and Paul Hopper (eds.)
2001 *Frequency and the Emergence of Linguistic Structure.* John Benjamins Publishing Company.
Bybee, Joan and Joanne Scheibman
1999 The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics* 37: 575–596.
Bod, Rens
1998 *Beyond grammar: An experience-based theory of language.* CA: Center for the Study of Language and Information.
Bukach, C.M., Isabelle Gauthier and Michael J. Tarr
2006 Beyond faces and modularity: the power of an expertise framework. *Trends in Cognitive Sciences* 10: 159–166.
Caldwell-Harris, Catherine L. and Alison L. Morris
2008 Fast Pairs: A visual word recognition paradigm for measuring entrenchment, top-down effects, and subjective phenomenology. *Journal of Consciousness and Cognition* 17: 1063–1081.
Carr, Thomas H.
1986 Perceiving visual language. In K.R. Boff, L. Kaufman and J.P. Thomas (eds.), *Handbook of perception and human performance.* New York: Wiley.
Davies, Mark
2010 Corpus of Contemporary American English COCA. Available from http://www.americancorpus.org/
Divjak, Dagmar
2008 On (in)frequency and (un)acceptability. In Barbara Lewandowska-Tomaszczyk (ed.), *Corpus Linguistics, Computer Tools and Applications – state of the art*, 213–233. Frankfurt: Peter Lang.
Donovan, John, J. and David J. Radosevich
1999 A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology* 84: 795–805.

Edelman, Shimon
2010 On look-ahead in language: navigating a multitude of familiar paths. In M. Bar (ed.), *Prediction in the Brain*. Oxford, UK: Oxford University Press.
Eder, Andreas B., Bernhard Hommel and Jan De Houwer
2007 How distinctive is affective processing? On the implications of using cognitive paradigms to study affect and emotion. *Cognition and Emotion* 21: 1137–1154.
Ellis, Nick C.
2002 Frequency effects in language processing. *Studies in Second Language Acquisition* 24: 143–188.
Ellis, Nick C. and Diane Larsen-Freeman
2009 *Language as a Complex Adaptive System*. Wiley-Blackwell.
Ellis, Nick. C. and Diane Larsen-Freeman
2006 Language emergence: Introduction to the special issue. *Applied Linguistics* 27: 558–559.
Ellis, Nick C. and Rita Simpson-Vlach
2009 Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory* 5: 61–78.
Elman, Jeffrey L.
1995 Language as a dynamical system. In Robert F. Port and T. van Gelder (Eds.), *Mind as Motion: Explorations in the Dynamics of Cognition*, 195–223. Cambridge, MA: MIT Press.
Gibbs, Raymond W. and Nandini P. Nayak
1989 Psycholinguistic studies on the syntactic behaviour of idioms. *Cognitive Psychology* 21: 100–138.
Gilquin, Gaetanelle and Stefan Th. Gries.
2009 Corpora and experimental methods: a state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5: 1–26.
Gries, Stefan Th.
2010 Dispersions and adjusted frequencies in corpora: further explorations. In Stefan Th. Gries, Stefanie Wulff, & Mark Davies (eds.), *Corpus linguistic applications: current studies, new directions*, 197–212. Amsterdam: Rodopi.
Hale, John
2003 The information conveyed by words in sentences. *Journal of Psycholinguistic Research* 32: 101–123.
Harris, Catherine L.
1998 Psycholinguistic studies of entrenchment. In J. Koenig (Ed.), *Conceptual Structures, Language and Discourse, Vol. 2*. Berkeley, CA: CSLI.
Jackendoff, Ray
1995 The boundaries of the lexicon. In M. Everaert, E.J. van der Linden, A. Schenk and R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives*. Hillsdale, NJ: Erlbaum.

Langacker, Ronald W.
1987 *Foundations of cognitive grammar. Vol. I: Theoretical prerequisites.* Stanford, CA: Stanford University Press.
Lewis, Richard L. and Shravan Vasishth
2005 An activation-based model of sentence processing as skilled memory. *Cognitive Science* 29: 375–419.
Lund, Kevin and Carl Burgess
1996 Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers* 28: 203–208.
MacWhinney, Brian
1991 *The CHILDES Project: Tools for analyzing talk.* Hillsdale, NJ: Erlbaum.
MacWhinney, Brian Ed.
1999 *The emergence of language.* Mahwah, NJ: Erlbaum.
McClelland, James L. and Kathryn Patterson
2002 'Words Or Rules' cannot exploit the regularity in exceptions: Reply to Pinker and Ullman. *Trends in Cognitive Sciences* 11: 464–465.
McClelland, James L. and Davd E, Rumelhart
1981 An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review* 88: 375–407.
Pinker, Steven
1999 *Words and rules: the ingredients of language.* New York: Basic Books.
Pustejovsky, James
1995 *The Generative Lexicon.* Cambridge, MA: MIT Press.
Ratcliff, Roger and Gail McKoon
1997 A counter model for implicit priming in perceptual word identification. *Psychological Review* 104: 319–343.
Stadthagen-Gonzalez, Hans, Jeffrey S. Bowers and Markus F. Damian.
2004 Age-of-acquisition effects in visual word recognition: evidence from expert vocabularies. *Cognition* 93: B11–B26.
Swinney, David and Anne Cutler
1979 The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior* 18: 523–534.
Solan, Zach, David Horn, Eytan Ruppin and Shimon Edelman
2005 Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences* 102: 11629–11634.
Tanenhaus, Mark K., Michael J. Spivey-Knowlton, Katherine M. Eberhard and Julie C. Sedivy
1995 Integration of visual and linguistic information in spoken language comprehension. *Science* 268: 1632–1634.

Tomasello, Michael
2003 *Constructing a language: a usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.
Waterfall, Heidi R., B. Sandbank, Luca Onnis and Shimon Edelman
2010 An empirical generative framework for computational modeling of language acquisition. *Journal of Child Language* 57: 671–703.
Wray, Alison
2002 *Formulaic language and the lexicon.* Cambridge, UK: Cambridge University Press.

Appendix: Data analyzed in word pair frequency study

Item		COCA (raw)	Google (Log)	Familiarity	Perceptual Identification
Random Pairs					
weep	job	0	3.85	0.32	0.29
butter	ace	0	4.06	0.5	0
comedy	span	0	4.08	0.91	0.29
pigs	troop	0	4.32	0.41	0
victim	cheese	0	4.62	0.14	0.25
course	hoop	0	5.2	0.64	0.29
basis	coast	0	5.32	0.27	0
cast	bark	0	5.47	0.18	0.415
blood	plane	0	5.87	0.27	0.165
pants	cloud	0	5.87	0.23	0.29
look	fever	0	5.98	0.32	0.29
desk	marks	0	6.14	2	0.25
cash	tone	0	6.19	0.45	0.415
belt	trade	0	6.29	0.77	0.875
taxi	tie	0	6.47	0.27	0.415
blast	brick	0	6.54	0.59	0.46
heart	root	0	6.58	0.18	0.25
smoke	bone	0	6.79	0.55	0.29
eyes	trees	0	7.01	0.32	0.415
art	beard	0	7.16	0.36	0
puppy	hill	0	7.22	0.45	0.25
anchor	stream	0	7.41	0.73	0.125
mass	floor	0	7.67	1	0.5
stroke	break	0	8	0.73	0.25
school	belly	0	8.48	0.18	0.25
hey	wing	0	8.68	0.18	0.25
trick	boy	0	9.38	0.59	1

dime	finger	0	9.68	0.18	0.29
pin	since	0	10.15	0.23	0.335
edit	center	0	10.52	1.14	0
days	group	0	11.66	0.86	0.58
wife	board	1	6.55	0.18	0.415
golf	where	1	11.21	0.68	0.25
home	leg	1	11.41	1.27	0.835
work	use	1	13.77	1.59	0.5
while	base	3	10.95	0.32	0.415
city	away	5	11.59	1.27	1
war	say	6	12.55	0.32	0.455
cold	off	8	10.81	0.41	0.5
system	never	28	12.9	0.5	0.25
Legal Pairs					
silly	trail	0	6.64	0.73	0.46
small	fuss	0	7.66	2.82	0.75
happy	name	0	10.09	1.14	0.33
tough	sort	1	6.35	2.05	0.165
proper	widow	1	8.16	1.23	0
dead	bride	1	9.91	1.27	0.29
caring	words	1	10.47	3.64	0.415
social	bunch	2	6.49	1.82	0.75
lousy	law	2	7.44	2.14	0.125
cruel	cost	2	8.52	1.32	0.46
famous	angel	2	9.5	1.36	0.75
modern	barn	2	9.57	0.86	0.25
simple	trend	2	10.59	2.14	0.71
right	unit	2	11.32	1.73	0.415
huge	church	3	10.99	3.32	0.5
left	step	3	11.17	1.91	0.415
early	change	4	11.13	1.77	1
true	people	5	13.11	2.64	0
mind	bomb	6	11.46	0.73	0.75
real	skin	6	11.86	1.41	0.58
same	run	6	12.69	1.5	0.625
entire	survey	7	12.28	3.36	0.375
sale	ends	7	14.59	4.05	0.71
size	three	8	12.29	3.18	0.54
public	attack	10	11.01	2.82	0.585
some	cans	13	11.46	2.18	0.75
gray	eye	14	10.41	1.86	0.54
empty	world	15	11.87	1.18	0.705
such	space	18	13.91	1.86	0.415
lost	girl	20	13.01	2.64	0.75
pay	rate	20	15	3.05	0.71

green	skirt	21	11.47	3.23	0.54
best	woman	23	13.04	1.95	0.415
new	table	23	14.43	3.27	1
open	spot	37	12.22	2.5	0.29
night	man	37	12.77	1.45	0.165
major	case	52	12.83	3	0.5
good	race	66	13.56	3.18	0.375
blue	wall	73	12.42	2.23	0.54
little	food	165	13.63	2.64	0.71
her	list	245	13.88	2.5	0.5
each	state	1001	17.21	2.82	0.585
two	ways	1726	17.9	3.73	0.71
both	men	2694	16.77	2.91	1
Collocations					
cents	worth	33	14.21	2.82	0.835
full	refund	46	16.46	4.27	0.415
local	bus	54	15.71	3.64	0.415
killer	bees	62	13.18	3.82	0.415
death	bed	66	13.9	4.18	0.585
sole	reason	72	14.03	3.09	0.455
hot	wheels	88	15.8	3.73	1
eight	ball	99	15.16	4	1
any	clues	101	14.19	2.95	0.25
litter	box	121	14.63	4.14	0.875
book	sales	136	16.11	3.77	0.5
high	esteem	140	14.06	3.68	1
money	order	147	18.03	4.27	1
pet	store	182	15.51	4.09	0.71
own	risk	184	17.64	4.14	0.75
mere	fact	242	15	3.82	0.75
safe	bet	266	14.55	3.68	0.585
blind	date	294	15.4	4.14	0.71
junk	mail	295	16.69	4.05	0.875
fan	club	310	17.07	4.14	1
wonders	why	339	14.87	3.36	0.415
die	hard	355	16.5	3.55	0.83
next	phase	407	16.24	3	0.125
top	secret	408	16.67	4.32	0.25
must	admit	613	16.56	3.59	0.75
zip	code	658	19.66	4.41	0.71
upper	hand	688	15.34	4.23	0.75
low	level	723	17.92	3.5	0.58
fair	trial	735	15.38	4	1
face	value	751	16.76	4.55	0.46
rush	hour	763	13.36	4.45	0.835

phone	lines	843	15.99	4.36	0.875
feel	free	927	19.17	4.68	0.335
peace	corps	928	16.49	4.05	0.83
deep	inside	973	15.98	3.82	0.835
screwed	up	978	16.37	4.32	0.75
machine	gun	1017	15.96	4.18	0.71
broad	range	1160	18.45	3.73	1
talk	shows	1360	16.39	4.55	0.875
focal	point	1451	17.33	3.77	0.705
back	yard	1533	16.33	4.36	0.75
keep	track	1632	18.1	4.27	0.71
brown	hair	1635	15.94	4.55	0.705
vice	versa	1895	18.19	4.5	0.835
news	media	1943	18.13	4.05	0.58
gold	medal	1983	16.77	4.14	0.75
black	hole	2048	17.26	3.82	0.705
front	page	2126	19.76	4.5	0.835
child	abuse	2156	17.21	4.64	1
locker	room	2424	16.38	4.27	0.75
post	office	2486	18.16	4.55	1
suffer	from	2704	17.82	3.64	0.75
get	married	3168	16.66	4.27	0.71
based	upon	3248	18.77	4.09	0.705
long	term	3400	20.15	4.36	0.75
very	nice	3536	18.25	4.68	0.875
natural	gas	3983	18.24	3.55	0.335
all	sorts	4459	18.01	4.05	0.665
credit	card	4932	19.95	4.55	0.75
few	hours	4938	17.81	4	1
shut	down	5403	18.05	4	1
much	better	6968	18.68	3.95	0.75
sounds	like	7340	18.51	4	1
months	ago	7962	18.48	3.82	0.585
parking	lot	8017	17.71	4.45	1
great	deal	10644	18.74	3.77	0.875
figure	out	12705	18.88	4.05	0.415
worry	about	12795	18.52	4.05	0.71
too	many	14153	19.1	4.55	1
middle	east	14780	20.13	4.32	0.585
five	years	24529	19.44	4	0.75
every	day	24947	19.62	4.68	1
health	care	28623	20.52	4.5	0.835
last	week	30436	19.57	4.5	0.5
rather	than	57700	20.39	4.32	0.665
thank	you	77530	20.45	4.77	1

Figurative extensions of word meaning: How do corpus data and intuition match up?

Jeannette Littlemore and Fiona MacArthur

Abstract

According to the cognitive linguistic paradigm, word senses sit within radial categories with basic, often concrete, senses at the centre and abstract, figurative senses lying towards the periphery. In this chapter we investigate the intuitions that both native and non-native speakers of English and Spanish had of the categories of senses associated with the words 'thread', 'wing', and *hilar* ('to thread') and compare these intuitions with findings from an earlier, corpus-based study of these words (Littlemore and MacArthur 2007), which had identified significant, yet different, sense shifts between equivalent denominal verbs in the two languages. Our results show that, compared with the corpus data, the intuitive data for both native and non-native speakers were relatively impoverished and skewed. Even advanced learners had limited knowledge of the senses lying towards the periphery compared with that of native speakers, and even among the native speakers there was considerable variation, with younger speakers exhibiting different knowledge from older speakers. We conclude that radial category knowledge builds up over a lifetime, and even rich corpus data is unlikely to reveal the variable nature of category knowledge among individuals.

1. Introduction and background to the study

In this chapter, we are interested in assessing the extent to which language corpora are able to predict a cognitive phenomenon. In other words, we seek to address a question largely ignored in corpus linguistics (but see Divjak 2010; Gries 2008; Taylor 2010), which is the relationship between the 'hard facts' of language as observable in large corpora and speakers' mental representation of language. The phenomenon that we are interested in here is that of radial categories. In cognitive linguistics, the various senses of individual words are thought to constitute radial categories, with the

most concrete, physical or 'basic' senses lying towards the centre, and the more abstract or figurative senses radiating out towards the edge (Taylor 2003). Polysemous senses of words can thus be charted as extending from a core meaning of a lexical item. The way these senses 'radiate' from the central category is, in cognitive linguistic parlance, 'motivated' in different ways, mainly through metaphor and metonymy. Radial categories are usually represented using a chart with a 'central' member (visually and semantically) from which other senses extend. Extensions usually follow sense shifts from more 'literal' senses to more abstract/figurative senses, which are represented visually as being furthest from the centre.

The literal to figurative continuum of word senses (Dirven 1985) becomes apparent if one examines the uses of a particular word in a large corpus. For example, if we look at the following citations from the Bank of English containing the word 'threaded', we can see that some of the senses of threaded such as those in citations (1), (4), and (6) are concrete and refer to actions involving pieces of thread or similar material, whereas others, such as those in citations (2), (3), (5), (7), and (9) are metaphorically extended to a type of movement that resembles threading. Citation (8) describes the positioning of the telegraph poles as a result of a threading process and citation (10) metaphorically maps the process of threading on to an abstract domain. Thus we can see in this small number of examples how the different senses of the word might be seen to radiate out from a core or basic sense via the processes of metaphor and metonymy.

(1) the nose of the spindle is **threaded** 20 by 2 mm, and there's a
(2) of every Saints attack — **threaded** a ball through the defence
(3) skipper Kavanagh, who **threaded** a low 18-yard shot past Posh
(4) Francis carefully **threaded** a microphone wire inside my jacket
(5) still evenings, as we **threaded** a path through narrow fjords
(6) returned to her room and **threaded** a needle she was making
(7) stood at my side. I **threaded** a way through surging traffic
(8) yachts and telegraph poles **threaded** along main streets, and
(9) the good-humoured protest **threaded** its way round Swindon,
(10) Slavery **threaded** its way as an issue, a concern,

It has been observed that those senses lying towards the edge of the category are more likely to form part of semi-fixed expressions. We can see this in citations (9) and (10) where the word 'threaded' forms part of the larger expression 'threaded its way'. This phenomenon can be viewed in different ways. It has been argued that phraseological units such as

these (whether continuous or discontinuous) provide a useful way of distinguishing between figurative and non-figurative uses (see Deignan 2005; Hanks 2006; Sinclair 1991). Construction grammarians, such as Goldberg (1995, 2006), would argue that in cases such as these it is the construction itself that carries the meaning and that the word 'threaded' is given this particular meaning by virtue of its presence within the construction, a process that Thornburg and Panther (1997) refer to as coercion. In our view, there is a two-way process involved here. Elements of the word 'threaded' allow it to fit well within this construction whereas the construction itself pulls the meaning of 'threaded' in a particular direction. All in all, this example serves to show how radial categories integrate flexibility with structural stability (Brone et al. 2006).

One of the aims of this chapter is to map the distribution of senses within radial categories using language corpora. Early accounts of radial categories tended to rest on introspection or the expert knowledge of word senses of the analyst (see, for example, Lakoff 1987, on the multiple senses of *over*). In recent years, usage-based accounts of polysemy have emerged which are based on the study of large corpora (see, for example, Gries (2006) on the polysemy of *run*). Although such accounts are data-based, they are not data-driven: the identification of the sense relations of polysemous words still relies largely on the analyst's intuitions or expert knowledge for a number of reasons. Among these are the fact that the 'motivation' of a sense extension is not something observable in linguistic data; another fact is that frequency counts of word senses in a synchronic corpus will not necessarily lead to a satisfactory identification of the core or basic sense of a word (for example, *right* used as a discourse marker will be found more frequently in most megacorpora than *right* signifying direction or the sense of 'correct'). The central member of the category may be historically older, and indeed more concrete than other senses found in the corpus, but it will not necessarily be the most frequent.

A second aim of the chapter is to compare corpus data with the radial category knowledge that informants are able to recall when given a single-word prompt. To the best of our knowledge there have been no studies comparing the radial category knowledge that exists in the minds of speakers with that which can be found in language corpora. Other areas of language knowledge that *are* starting to be investigated in this way include frequency of lemma types (Taylor 2010), word class (this volume) and noun combinations (this volume). Findings from these studies indicate that there are parallels between the intuitive knowledge that people have

of the distribution patterns that exist within languages, and the patterns that can be identified in language corpora. Taylor (2010) demonstrates for a number of areas how an individual's mental grammar (or 'I' language) builds up through exposure to multiple exemplars and the use of probabilistic processing and 'intuitive statistics'. Through these processes, we are able to extrapolate principles of a mental grammar from the language to which we are exposed. However, despite these initial forays, very little is known about the way in which people's intuitive knowledge of *radial categories* reflects the patterns that can be found in language corpora. One might expect the sense relations that are available in corpora to be reflected in the speaker's mental grammar. One might even hypothesise that there is something like a 'mental corpus', i.e. a usage based system, reflecting all the usage events to which we have been exposed, which presumably will also be similar to those found in a large corpus.

However, the reality is unlikely to be as straightforward as this. Knowledge of radial categories is built up over a lifetime of exposure to language and some senses (and their corresponding grammar patterns) are likely to be more salient than others for different speakers at different times (Giora 2003). Although in general terms we agree that Sinclair's (1991) distrust of intuition as a reliable guide to usage is well-founded (among other reasons because there are factors that affect people's ability to recall language knowledge), it is important not to reject intuitive data completely, but rather to find mutual enrichment in the relationship between corpus data and intuitive data. One would expect there to be certain features of language that are revealed by corpus data but which tend to be less salient in the minds of native speakers. Equally, one might expect speakers of a language to be aware of certain features of language (in particular those that involve word meaning and relations between senses) that are not revealed by language corpora. By identifying these areas, it is easier to see how corpus and intuitive data might complement one another, thus leading to a better understanding of language itself and of the psycholinguistic processes that underlie it. For example, on the psycholinguistic front, it is useful to examine the factors that trigger recall and what factors lead to the suppression of knowledge. In other words, it would be useful to know what it is that makes some types of linguistic knowledge more salient than others, a fact that is likely to account to some extent for the differences between corpus data and intuitive data. Such differences may well be important for language teaching, as corpus data alone may very well be misleading if taken to represent the I-language of a native speaker

and thus not provide a fully reliable guide to what should be the target of the non-native speaker in the quest to reach native-speaker-like competence in the L2.

In this study, we investigate what radial category knowledge speakers are able to recall when prompted, and compare it to that which is revealed in corpus data. The focus is on relatively polysemous denominal verbs. In order to increase the reliability of our study it was conducted in two languages: English and Spanish. Radial categories of word senses have been found to develop in different ways in different languages and one would expect speakers of those languages to have knowledge of these respective categories. We therefore took two more or less equivalent verbs in these two languages: English 'to thread' and Spanish *hilar*, and English 'to wing' and Spanish *aletear*. In a previous study (Littlemore and MacArthur 2007) we studied both of these verbs in depth using the *Bank of English* (BofE) and the *Corpus de Referencia del Español Actual* (CREA) and found that there were differences in terms of the ways in which they could be figuratively extended in English and in Spanish. In the present study, we are particularly interested in looking at our informants' conscious knowledge of word senses, as *conscious* knowledge was the focus of Sinclair's earlier criticisms.

Unlike the results for 'thread', *hilar*, and 'wing', a pilot study revealed that, when Spanish-speaking informants were asked about the senses they associated with the verb *aletear*, none were able to come up with more than one sense for this verb (the basic sense of 'flap' of a birds' wings), despite its polysemy and relative frequency in the corpus. We therefore decided not to continue exploring speakers' intuitions about the senses of this verb as this initial study revealed that speakers' response to its polysemy were so impoverished that little would be gained from further exploration using the method described below. The Spanish intuitive data thus deal solely with the verb *hilar*. A pilot study with the English-speaking participants revealed that they found it impossible to focus solely on the verb forms of 'thread' and 'wing' as they automatically produced responses involving the noun form. We therefore decided to focus on both forms in the English part of the study. For the purposes of comparison, a further corpus study was conducted to identify the radial network of the noun (and in rare cases, adjectival) forms of these words in the Bank of English.

The study also looks at the radial category knowledge that second language learners have for these words. Research has shown that even advanced language learners are very often unaware of the figurative exten-

sions of word meanings (Alejo 2008) so we were interested to see how their category knowledge maps onto the type of information that can be found in the language corpus, and how it compares to that of native speakers. Our research questions were therefore as follows:

1. In what ways do English and Spanish differ in terms of the ways in which the denominal verbs *thread*/*hilar* and *wing*/*aletear* are extended?
2. For each language, how do corpus-data compare with intuition?
3. What factors influence radial category knowledge and how is it retrieved?
4. In what ways do language learners differ from native speakers in terms of their radial category knowledge?

2. Methodology

2.1. The corpora

In order to answer the first research question we used the 450 million-word BofE and the 75 million word Peninsular Spanish section of the CREA. The language data contained in these two synchronic corpora are regarded as sufficiently representative to be used to inform much linguistic description, as well as lexicographical work or even language teaching materials.

2.2. The participants

In order to answer Research Questions 2 and 3 we interviewed 20 native speakers of English, 26 native speakers of Spanish, 20 non-native speakers of English with mixed L1 backgrounds (all of whom were upper-intermediate/advanced and living in the UK) and 9 non-native speakers of Spanish with mixed L1 backgrounds (all of whom were upper-intermediate/advanced and living in Spain). Individual interviews were conducted with informants, and these were taped and later transcribed. In line with our respective universities' research ethics policies, the informants were told that we would transcribe the interviews and that any data used would be anonymised. They were given the prompts: 'I'm going to ask you to give me all the senses or meanings you can think of for the words 'thread' and 'wing' You can use it in any form or part of speech', in the case of the English words (we verified that the informants knew what was meant by

'part of speech' asking them for examples relating to everyday vocabulary prompts) and *Que significados y usos conoces del verbo 'hilar'?* ('What senses/meanings or uses do you know of 'to thread'?) in the case of the Spanish words. The interviewers checked that the prompts had been fully understood before encouraging the informants to think of as many senses as possible. The responses were analysed quantitatively and qualitatively.

We collected demographic data from each of our participants. The native-speaking informants were not a uniform group, but varied in terms of age, gender, and educational background. The age of the native English-speaking informants on the English verbs, nouns and adjectives ranged from 20 to 51; 12 were female and 8 were male. The age of the non-native informants on the English verbs, nouns, and adjectives ranged from 20 to 35; 16 were female and 4 were male. The age of the native Spanish-speaking informants on the Spanish verb ranged from 19 to 58 years old; 14 were female and 13 male. Of the 9 non-native-speaking informants for this verb, the age range was 21 to 46; 5 were female, 4 male. Where relevant, differences among informants' responses are discussed in relation to these variables, and the terms 'expert' and 'naïve' will be used to distinguish between older (45–60 years old) informants trained in linguistic description and other speakers with little or no formal knowledge of linguistics respectively. The non-native speakers in the study had a range of first languages (L1s) including Chinese, Polish, German, Italian, and French.

3. Results

The results of this study are presented in order of the research questions.

3.1. Research Question 1: In what ways do English and Spanish differ in terms of the ways in which the denominal verbs 'thread'/'hilar' and 'wing'/'aletear' are extended?

Littlemore and MacArthur (2007) give an account of the sense relations of the polysemous verbs *thread*, *hilar*, *wing*, and *aletear*, which are set out schematically in the tree diagrams in Figures 1–4. In order to identify the different senses of these verbs, we carried out collocation and colligation analyses of the usage events involving these verbs recorded in the two corpora mentioned. That is, we examined the lexical and syntactic co-selection of items surrounding the node words. This allowed us to separate

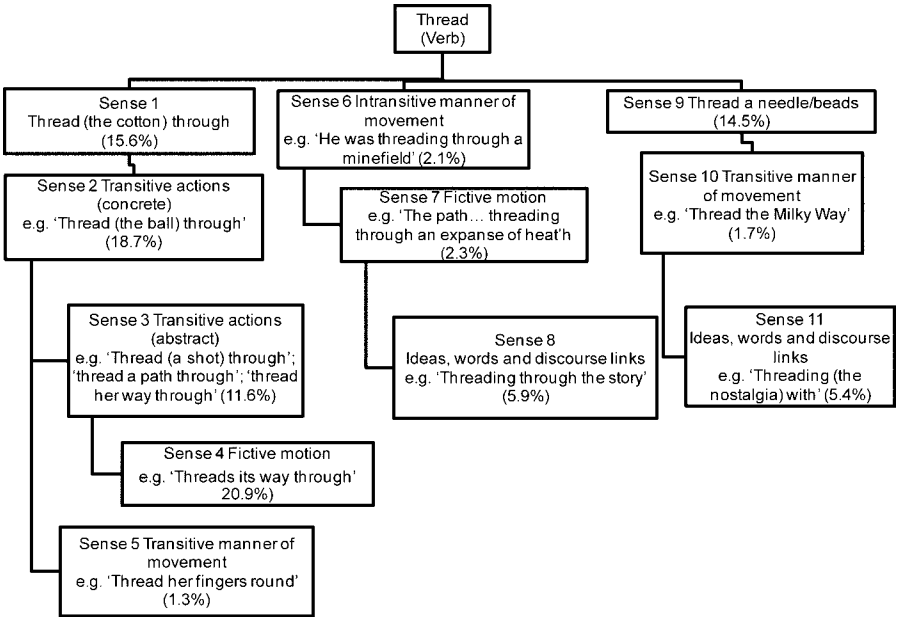


Figure 1. Thread (verb) 717 citations (1.59 per million words) Includes search items: 'to thread'; 'threaded'; 'threading', 'thread/VERB, and 'threads/VERB' (erroneous hits removed)

transitive from intransitive uses of the verbs, and to identify, for example, movement senses of 'wing' and 'thread' through the presence of prepositional phrases or locational complements to the right of the verb or its direct object. The 'behavioural profile' of each verb (Hanks 1996, 2006) revealed by this type of analysis meant that the first of the three criteria recommended by Evans (2005: 41) in his plea for a 'principled approach' to polysemy – that a distinct sense must contain additional meaning compared to other already established senses – was not always relevant. For example, one figurative sense of *hilar* might be glossed as 'join' or 'link', but an analysis of the collocates to the right of the verb reveals that not any entity may be figuratively linked using the verb *hilar*. Rather, typical collocates included ideas, actions, words (or larger linguistic units), revealing 'semantic preferences' (Sinclair 1991). These details are likely to be useful to language learners but are lost if the analyst decides to conflate these usage events into one sense. Likewise, we were also able to associate senses with particular patterns (e.g. verb aspect, restricted collocations).

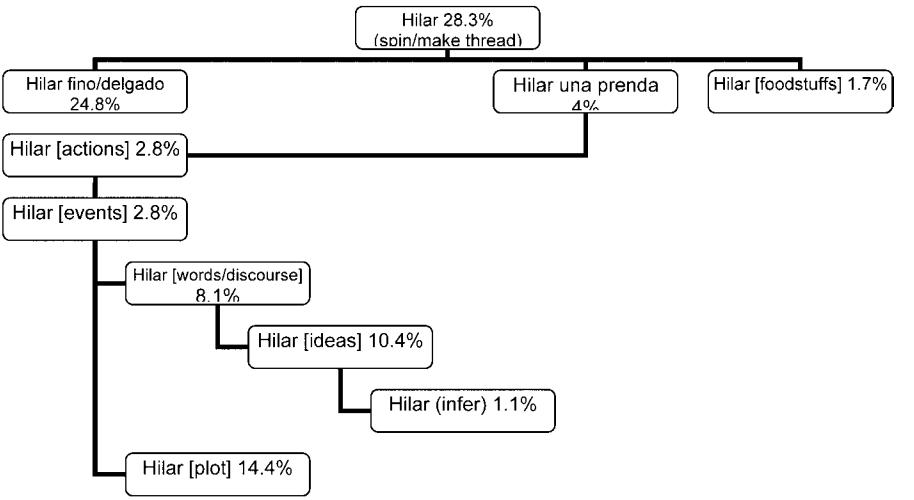


Figure 2. Hilar (overall frequency in the CREA 2.38 p/M). Search items included all forms of the verb HILAR (frequencies rounded to the nearest percentage)

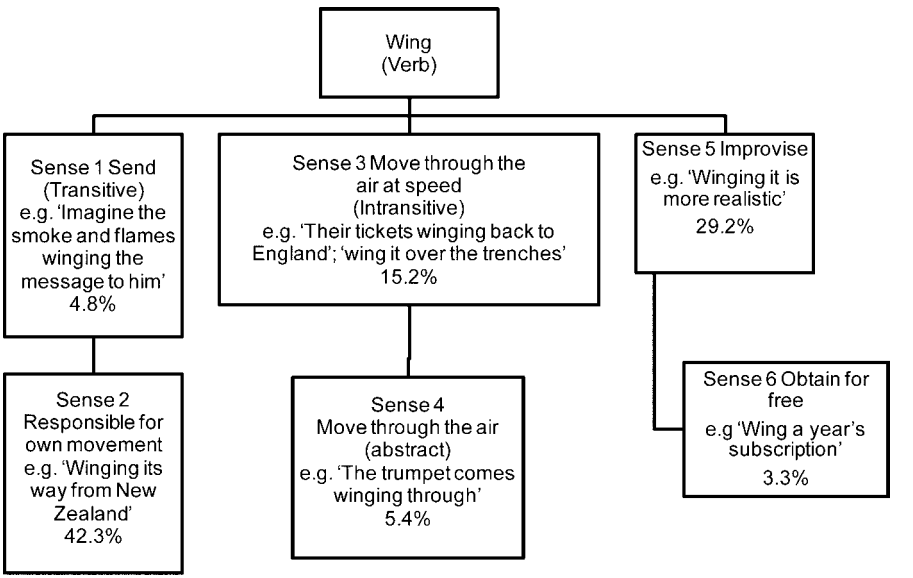


Figure 3. Wing (Verb) 336 citations (0.74 per million words) Includes 'to + wing'; 'winging'; 'winged'; wing/VERB, and wings/VERB (erroneous hits removed)

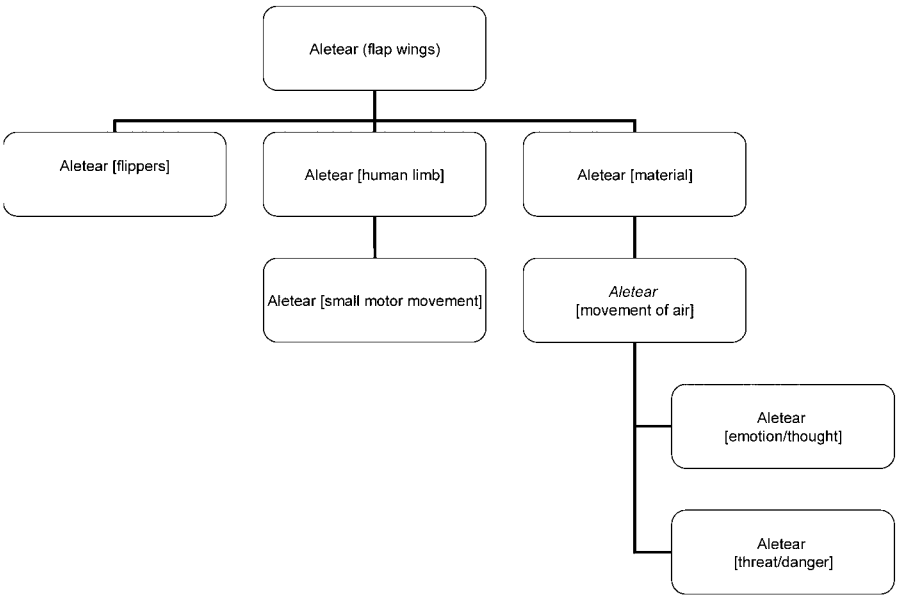


Figure 4. Aletear (overall frequency in CREA 1.5 p/M). Search items included all forms of the verb ALETEAR

That is, our original research questions and our focus on the foreign language learner led to an analysis that was more fine-grained and detailed than might be considered appropriate or even desirable by other researchers looking at polysemous words such as these.

The senses thus identified were grouped according to their semantic relation with each other, and represented schematically in a manually-constructed tree diagram. For instance, the transitive uses of 'thread' in the pattern Subject + THREAD + Direct Object + Locational Complement were separated from the intransitive uses of the verb in the pattern Subject + THREAD + Locational Complement and were represented as separate 'branches' on the tree diagram. The senses were then placed on the branch following a cline from most concrete (at the top) to most abstract or figurative (at the bottom), this cline being the result of the analysts' intuitions based on the corpus evidence. So, for example, *hilar una prenda* (to join/sew an item of clothing) involves a direct object referring to a concrete entity and is therefore placed at the highest point on the branch grouping transitive uses of this verb; *hilar una idea* (to link an idea), in contrast, is metaphorical as it involves a direct object that refers

to an abstract concept, and is accordingly placed lower or further away on the same branch. For ease of reference to these senses in the written text, rather than with reference to their relative importance, the senses were assigned a number on the diagram.

An important difference between the tree diagrams schematically representing the multiple senses of these verbs is the presence or absence of a core sense for each verb. In the case of the Spanish verbs, the 'central' members are placed at the top centre of the diagrams. This core or 'sanctioning' sense (Evans 2005) tended to be more concrete and historically older than the other senses (and thus seen to 'motivate' the figurative extensions of the word). In those cases where more than one concrete sense was associated with a word (as found with *hilar*), frequency in the corpus decided its relative status in relation to other senses. That is, in the case of *hilar*, the most frequent concrete sense was 'spin' rather than 'sew', and it was this sense that was assigned the status of 'core' sense (a decision in line with that of dictionaries such as the 22nd Edition of the *Diccionario de la Real Academia Española* (2001), which does not even record the 'sew' sense).

In the case of the English verbs, the central member for the radial categories is not represented by a core sense of the verb, but rather a 'word': *thread* and *wing*. This representation is designed to show the analysts' lack of commitment to the notion that verbs and nouns with identical form can or should be separated in terms of their sense relations.[1] As will be seen, this decision also reflects the performance of the informants in the part of the study designed to collect the intuitive data. These informants moved backwards and forwards between the different parts of speech and appeared to be oblivious to them. Unlike the Spanish verbs, which are the products of derivational processes that sets them apart from the nominals *hilo* and *aleta*, the verbal senses of *thread* and *wing* have come about through a process of grammatical shift, and charting the complete radial category for these words would most likely have involved representing the senses associated with the nouns on the same chart, with a meaning associated with the noun as the core or sanctioning sense for both verbal and nominal uses. In this paper the nouns and verbs are shown in separate charts in order to facilitate comparisons with the Spanish data (which, for the reasons outlined above, only focused on the denominal verbs). However, a more accurate reflection of the radial categories for English (in both the corpus and in the informants' data) would involve a single chart combin-

1. For further discussion of this point, see Tranel et al. (2005) or Tyler et al. (2008).

ing the different word forms. In short, although the analysis offered is of denominal verbs in two languages, the formation of the verbal forms was also seen to be relevant in charting their respective sense relations.

Unlike other studies of word meanings such as that of Xiao et al. (2009), our analysis did not provide any information about the dispersion of the senses identified throughout the corpus; that is, we did not formally chart the relationship between the different senses found and the kinds of texts in which they were represented. However, as will be seen below, we noted informally that certain senses tended to be associated with particular topics and genres.

In Figure 1 we can see that there were three branches of senses for 'thread'. The middle branch contains intransitive senses whereas the left and right branches contain transitive senses that are different from one another, yet related metonymically (i.e. 'thread a needle' has a metonymic relationship to 'thread the cotton through the needle'). Each of these three branches then extended through a series of sense changes involving manner of motion, fictive motion and concrete to abstract changes.

The three branches of senses of *hilar* identified in the corpus are seen to be differently related to the transitive and intransitive uses of the 'spin' sense of the verb. Two of these branches contain only one item, both of which are semi-fixed expressions. The semantically most 'productive' branch consists of a number of senses related to a metonymically motivated shift in the core sense of the verb (the product of spinning, namely the thread, stands for the action of sewing), which extends its meaning to a general sense of linking and joining. Here they are charted according to the types of direct objects found to the right of the verb, with the most abstract appearing furthest away from the concrete sense that sanctions them.

In Figure 3 we can see that there were three branches of senses for 'wing', one transitive, one intransitive, and a third branch of senses that were markedly different from, yet metaphorically related to the other two branches. ('winging it', meaning 'to improvise' was different from, yet metaphorically related to the other senses of 'winging' which contained elements of movement and directionality). The senses then developed along these branches by becoming more reflexive or more abstract. In the third branch, there was a loose relationship between 'improvisation' and 'obtaining something for free' (possibly relying on underlying notions of minor criminality).

Aletear derives from *aleta* (fin) but the verb is also used to refer to locomotion with the use of the wings: of these two senses of self-propelled

motion through air and water, air-borne movement is the most frequent in the corpus and motivates a number of related senses, visually represented in the branch on the right. These range from uses of the verb to literally describe the movement of light materials in a breeze to metaphoric senses of the verb to denote cognitive and affective processes. The senses associated with the different limbs that enable locomotion (fins and wings) are separated from each other and separated also from the use of the verb to describe movements of parts of the human body (the arms or nostrils, for example), which do not imply locomotion but rather focus on a flapping motion.

Despite choosing denominal verbs whose nominal forms refer to exactly the same ('thread' and '*hilo*') or similar ('wing' and '*aleta*') real-world entities, the verbal forms display a number of cross-linguistic differences. There were, however, some similarities across the two languages. For example, the complementation patterns of these verbs reveal significant sense shifts; all the verbs are found used transitively and intransitively in the two corpora. Sense shifts are associated not only with the different complementation patterns of the four verbs, but also with particular fixed patterns (*thread* found in the 'way-construction', for example, or *hilar* in the idiom *hilar fino*). However, there was remarkably little overlap between the sense extensions of the target items, except in the linking/joining senses of *thread* and *hilar*. The radial categories themselves were more or less 'elaborate' in the two languages. *Thread* has a more complex network of senses in comparison with *hilar*; but *aletear* (based on corpus evidence) has a more complex radial category than *wing*. The overlap between some of the sense extensions shared by *thread* and *hilar* can probably be explained as arising from the impact of Latin on English through literal translation of classical authors in Renaissance and after (Spanish and English both 'inherit' the figurative extensions of *thread* and *hilar* meaning 'join' from Latin).The different ways in which Latin affected the two languages is evident in the relative frequency of these senses in the two corpora; the joining sense of 'thread' appears to be used far less frequently in English than it is in Spanish. Otherwise, the ways in which the senses extend are different in that *thread* focuses more on careful, precise or wavy-shaped movement whereas *hilar* focuses more on joining or linking. *Wing* conveys a range of senses related to improvisation and transport which it does not have in Spanish. *Aletear* is used more to refer to flapping movements and extends to describe the movement of the air and emotions (see Littlemore and MacArthur 2007 for details).

3.2. Research Questions 2 and 3: For each language, how do corpus-data compare with intuition? What factors influence radial category knowledge and how is it retrieved?

In order to provide a fuller picture of our findings we present the findings for these two research questions together. Research Question 2 requires a quantitative analysis of the data, whereas Research Question 3 involves a qualitative analysis of the possible reasons for these findings. In order to assess how the corpus data compared with the intuitive data, we calculated the proportions of responses given by the informants and mapped them onto the tree diagrams produced for the corpus data. A summary of the quantitative results is given in Tables 1–5.

3.2.1. Corpus data and intuitions compared: Quantitative analysis

We can see from Tables 1–5 that there was a strong tendency for the informants to mention the basic sense and not the figurative extensions of the English words. Interestingly, there was very little mention of the sports sense of either *thread* or *hilar*, which appear to be somewhat over-represented in both corpora. The most arresting difference with regard to *hilar* was in informants' mentions of the concrete senses of the verb, with very few referring to the core sense 'spin'. This is not surprising if one considers that fewer people in the twenty-first century will have real world knowledge of spinning than they will of its product and uses. However, the corpus data provided a very different picture of the status of the concrete senses of the verb (see also Sandra and Rice 1995, on spatial and temporal uses of English prepositions) and may partially explain the low number of mentions of the idiom *hilar fino* ('to do something carefully with great attention to detail') by respondents, for many of whom the phrase might be semantically opaque and non-compositional. Like its variant form, *hilar delgado*, the figurative motivation of the idiom would be obscure if the concrete sense of *hilar* is, for most informants, 'to sew/tack material together'. The intuitive data thus do not match up with the corpus data. The most frequent items in the intuitive data for English were the basic sense, the abstract reference to continuity and 'email threads', which were, in relative terms, less widely represented in the corpus. At first sight, the proportion of informants who produced the term 'thread-bare' also appears to be relatively high in comparison with the corpus data, however this difference is skewed by the methodology, which did not include the search term 'threadbare' as a separate item.

As pointed out in Section 1, a pilot study revealed that when informants were prompted to think about the English words 'thread' and 'wing',

although specifically asked to think about the verb, they often automatically produced the noun form. For this reason, in the main study, the participants were told that they could use any part of speech. Our results showed that the native speakers produced far more senses for the nouns than they did for the verbs (see Tables 1 and 2). The greater preference for the noun in this recall task is likely to be largely attributable to frequency effects (both thread and wing appeared in the nominal form far more frequently than in the verbal form in the corpus, as we can see in Tables 1, 2, 4 and 5). The finding may also be related to the greater ease with which nouns are processed in comparison to verbs (Spenney and Haynes 1989; Kauschke and Stenneken 2008) or the recall advantage for nouns that has been shown to be particularly marked for adults (Earles and Kersten 2000). Although the issue of how grammatical category information is represented in the mind/brain remains a contentious issue, the fact that nouns tend to encode entities while verbs express relational processes (Langacker 1987) may have made it more likely that entities would be recalled: imagery encoding might have proved easier in response to a decontextualised prompt such as that used here than motor encoding. Furthermore, the fact that the items asked about were denominal verbs (and hence in English, indistinguishable from the nouns in the prompt given) might have propitiated a focus on figurative extensions of the noun, for, as Goatly has pointed out (1997: 82–92) nominal realisations of metaphor appear to be especially marked or memorable. In the case of Spanish, however, no such form-class ambiguity is found, for all lexical words are obligatorily marked for grammatical category, as Spanish has a much richer inflectional system than English. Nevertheless, the Spanish-speaking informants also mentioned the noun *hilo*, as the following extract from the transcript shows, where there is specific mention of a mental image in relation to the figurative motivation of the 'linking' sense of *hilar*:

Extract 1
Bueno, hilar creo que principalmente se usa en el sentido de coser, de lo relacionado con un ***hilo****. Pero lo usamos más a menudo yo creo para hilar una historia o un argumento para que tenga sentido, que siga un* ***hilo, como un- una línea****.*

[Well, I think to spin is usually used in the sense of to sew, everything related to a **thread**. But I think we use it mostly to thread a story or a plot so that it makes sense, so that it follows **a thread, like a line**][2]

2. The translation offered here and of the other extracts loosely render the speaker's words, using 'spin' to translate 'hilar' when used intransitively and 'thread' when used transitively.

Table 1. Senses of 'thread' (verb): tokens and overall percentages of tokens in the Bank of English; mentions and overall percentages of mentions by native (NSs) and non-native speakers (NNSs)

Sense/form	Tokens in BoE (717 in total)	Overall percentage in BoE	Mentions by NSs (26 in total)	Overall percentage of mentions by NSs	Mentions by NNSs (5 in total)	Overall percentage of mentions by NNSs
1. v.t. (e.g. *thread* [*cotton*] *through*)	112	15.6%	10	38.4%	1	20%
2. v.t. concrete actions (e.g. *thread* [*the ball*] *through*)	134	18.7%	1	3.8%	0	0%
3. v.t. abstract (e.g. *thread* [*a path*] *through*)	83	11.6%	1	3.8%	0	0%
4. v.t. fictive motion (e.g. *threads* [*its way*] *through*)	150	20.9%	5	19.2%	0	0%
5. v.t. manner of movement (e.g. *thread* [*her fingers round*])	9	1.3%	0	0%	0	0%
6. v.i. manner of movement (e.g. *thread* [*through a minefield*])	15	2.1%	3	11.5%	1	20%
7. v.i. fictive motion (e.g. *the path threading through* [*a minefield*])	17	2.3%	0	0%	0	0%
8. v.i. ideas, words discourse (e.g. *threading through* [*the story*])	42	5.9%	0	0%	0	0%
9. v.t. (e.g. *thread a needle/beads*)	104	14.5%	6	23%	3	60%
10. v.t. manner of movement (e.g. *thread* [*the Milky Way*])	12	1.7%	0	0%	0	0%
11. v.t. ideas, words and discourse links (e.g. *threading* [*the nostalgia*] *with;* [*states*] *threaded together* [*by stratagem*])	39	5.4%	0	0%	0	0
Total	717	100%	26	100%	5	100%

Table 2. Senses of 'thread' (noun or adjective): tokens and overall percentages of tokens in the Bank of English; mentions and overall percentages of mentions by native (NSs) and non-native speakers (NNSs)

Sense/form	**Tokens in BoE (2442 in total) Figures below based on a sample of 500***	**Overall percentage in BoE**	**Mentions by NSs (41 in total)**	**Overall percentage of mentions by NSs**	**Mentions by NNSs (11 in total)**	**Overall percentage of mentions by NNSs**
1. n. Basic sense	205	41.1%	20	48.8%	7	72.8%
2. n. Only just holding on (physical sense) (e.g. *the limb was hanging by a thread*)	11	2.2%	0	0%	0	0%
3. n. Only just holding on (abstract sense) (e.g. *the dream was left hanging by a thread*)	44	8.8%	0	0%	0	0%
4. n. Linking/continuity (abstract) (e.g. *A thread of ideas*)	154	30.8%	12	29.2%	1	9.1%
5. n. A thread on a website	5	1.0%	3	7.3%	2	18.1%
6. n. A small amount of something concrete resembling a thread	45	9.0%	0	0%	0	0%
7. n. A small amount of something abstract resembling a thread	31	6.2%	0	0%	0	0%
8. adj. Thread-bare (e.g. *A thread-bare quilt*)	3	0.6%	6	14.6%	0	0%
9. adj. Thread-bare (abstract) (e.g. *A thread-bare laugh*)	2	0.4%	0	0%	0	0%
Total	500	100%	41	100%	11	100%

Table 3. Senses of 'HILAR': tokens and overall percentages of tokens in the CREA; mentions and overall percentages of mentions by native (NSs) and non-native speakers (NNSs)

Sense/form	Tokens in CREA (179 in total)	Overall percentage in CREA	Mentions by NSs (60 in total)	Overall percentage of mentions by NSs	Mentions by NNSs (18 in total)	Overall percentage of mentions by NNSs
1. v.i. spin (e.g. *la lana . . . se hilaba en la rueca*)	49	27.3%	5	8.3%	1	5.5%
2. v.i. *hilar fino/delgado*	43	24%	5	8.3%	1	5.5%
3. v.t sew/tack together (e.g. *hilando unos pantalones*)	6	3.3%	12	20.%	4	22.2%
4. v.t. actions (e.g. *su incapacidad de hilar jugadas*)	7	3.9%	0	0%	0	0
5. v.t. events (e.g. *una singular fata que hila el sino lastrado*)	5	2.7%	6	10.%	0	0
6. v.t. plot (e.g. *en ocho horas tenía una historia bien hilada*)	25	13.9%	6	10.%	3	16.6%
7. v.t. words/discourse (e.g. *no acierta a hilar las palabras*)	14	7.8%	12	20.%	5	27.7%
8. v.t. ideas (e.g. *incapaces de hilar correctamente un pensamiento*)	18	10%	13	21.6%	4	22.2%
9 v.t. infer (e.g. *La gente está hilando la separación con un romance con Melanie Griffith*)	2	1.1%	0	0%	0	0
10. v.t. (-en participle) threadlike appearance (e.g. *fiambre relleno de huevo hilado*)	10	5.5%	1	1.6%	0	0

Table 4. Senses of 'wing' (verb): tokens and overall percentages of tokens in the Bank of English; mentions and overall percentages of mentions by native (NSs) and non-native speakers (NNSs)

Sense/form	**Tokens in BoE (336 in total)**	**Overall percentage in BoE**	**Mentions by NSs (18 in total)**	**Overall percentage of mentions by NSs**	**Mentions by NNSs (0 in total)**	**Overall percentage of mentions by NNSs**
1. v.t. Send (e.g. *Imagine the smoke and flames winging the message*)	16	4.8%	0	0%	0	0%
2. v. reflexive To be responsible for own movement (e.g. *Winging its way from New Zealand*)	142	42.3%	4	22.2%	0	0%
3. v.i. Move through the air at speed	51	15.2%	0	0%	0	0%
4. v.i. Move through the air (abstract)	18	5.4%	0	0%	0	0%
5. v.i. Improvisation (e.g. *Winging it is more realistic*)	98	29.2%	14	77.8%	0	0%
6. v.t. Obtain for free (e.g. *Wing a year's subscription*)	11	3.3%	0	0%	0	0%
Total	336	100%	18	100%	0	100%

Table 5. Senses of 'wing' (noun or adjective): tokens and overall percentages of tokens in the Bank of English; mentions and overall percentages of mentions by native (NSs) and non-native speakers (NNSs)

Sense/form	**Tokens in BoE (26,147 in total) Figures below based on a sample of 500**	**Overall percentage in BoE**	**Mentions by NSs (51 in total)**	**Overall percentage of mentions by NSs**	**Mentions by NNSs (23 in total)**	**Overall percentage of mentions by NNSs**
1. n. Basic sense (bird or plane wing)	54	10.8%	13	23%	15	65.2%
2. n. Military (e.g. *He's got his wings*)	4	0.8%	4	7.8%	0	0%
3. n. Part of a building or other physical entity	36	7.2%	14	27.5%	4	17.4%
4. n. Part of an organisation or group (e.g. *The European wing*)	37	7.4%	0	0%	0	0%
5. adj. Political position (e.g. *Left-wing/ right-wing*)	282	56.4%	4	7.8%	3	13.0%
6. n. Improvisation (e.g. *A wing and a prayer*)	2	0.4%	10	19.6%	0	0%
7. n. or adj. Sports position (e.g. *wing-back*)	67	13.4%	5	9.8%	1	4.3%
8. n. or adj. Resembles a wing (e.g. *wing nuts*)	7	1.4%	1	1.9%	0	0%
9. n. Protection (e.g. *She took me under her wing*)	11	2.2%	0	0%	0	0%
Total	500	100%	51	100%	23	100%

This finding is of interest, if only because, although certain approaches to metaphor identification (for example, Steen et al. 2010) may well treat the linguistic products of derivational processes or grammatical shift as unrelated semantically, informants' responses showed that the sense relations between nouns and derived verbs are apparent to them.

The rich verbal paradigm for languages such as Spanish has been shown to have consequences for word processing, with differential activation for nouns and verbs (Perani et al. 1999; Longe et al. 2007), and speakers of Spanish will be interpreting inflectionally varying forms more often than speakers of a language such as English. In this regard, it was interesting to note that although the prompt form of the verb *hilar* in Spanish consisted of the stem and an infinitive marker, this did not appear to have any strong inhibitory effects on the responses given. For example, just over half of the tokens (55.8%) of *hilar fino* [to do things carefully] in the CREA were realised by the infinitive, and, of all infinitive uses of HILAR in the corpus, 41% were associated with the use of this idiom. The relative frequency of *hilar* in infinitive form in this idiom arises from its axiology: it was found to be preceded by verbs expressing obligation (*tener que*/*hay que*/*obligar a* [have to/be obliged to]) or others (*saber*, *aprender a* [know, learn to]) that likewise express the desirability of acting with care or subtlety in seeking to achieve some goal. The frequency of this form-meaning pair might have been expected to be reflected in informants' recall of the idiom in response to the infinitive form of the verb as a prompt. This was not the case, however, as only five informants mentioned the idiom. One possible reason for this might be found in speakers' mental representation of the radial category for this word. As has been noted, the core sense of *hilar* is 'to spin', a sense only mentioned by a small number (or 8.3%) of informants. From the analyst's point of view, this sense is seen to sanction or motivate the metonymic and metaphoric extensions of *hilar* found in the corpus data, and this idiom most directly. However, if the 'spin' sense did not enjoy this privileged status in a language user's radial category of senses for the word, and the transitive use of the verb (meaning 'sew' or 'tack together') was the sense that sanctioned the figurative extensions of the verb, the idiom *hilar fino* might be then be perceived as being unrelated semantically to other uses of the verb mentioned, which suggests that these speakers might process it as an unanalysed unit. The canonical idiom or 'metaphoreme' (Cameron and Deignan 2006) did not appear to form part of the radial category of senses for *hilar* for most of these informants, which agains raises the issue of the representativeness of language corpora. Pure (classical) idioms are rarely

used in their canonical form (Moon 1998) and the behaviour of the informants may have simply reflected this fact. Canonical idioms may be slightly over-represented in the CREA. This corpus is somewhat dominated by books, periodicals, and magazines, and it may be that these genres are more likely to contain pure idioms than other genres.

Although the informants tended not to produce canonical idioms, they did appear to be sensitive to the more subtle phenomenon of restricted collocations. In fact, the presence of restricted collocations with or without alliteration appeared to aid recall for all participants, but most particularly the English informants. Many of the participants were able to recall 'wing it' before they produced the basic sense of 'wing'. When they went back to the basic sense they then appeared to begin a new search for meaning, as we can see in Extract 2:

Extract 2

Interviewer: And the second word is 'wing'

Participant: Wing it. Aeroplane wing, bird's wing, winging it erm winging your way, merrily on your way erm, I'm trying to wing it now. Winging your way through a dance erm, yes to wing something, wing on a plane, bird's wing erm football you have a wing player plays on the wing of a football ground erm yes you wing your way through something, winging it, erm I think seeds have wings don't they? That's how they get to the ground erm what has wings? I'm thinking of things that have wings.

Participants seemed to be sensitive to the fact that certain parts of speech are more likely to be used with certain fixed expressions. For example, 'threading' was produced more often in the context of 'threading your way through' than in any other context.

On the other hand, the informants were not constrained by the verbal form of the prompt, as just one extract from the transcripts illustrates when a restricted collocation (*huevo hilado* ['spun' egg]) was the first sense recalled:

Extract 3

Interviewer: Entonces la pregunta es sobre el verbo hilar.

Participant: Hilar

Interviewer: Hilar

Participant: Huevo hilado. Eh, eh, hilar una conversación y-y-y. en el sentido de trenzar, eh, en ese sentido

[Interviewer: So the question is about the verb to spin.

Participant: To spin

Interviewer: To spin

Participant: Spun egg. Um, to thread a conversation, and, and, in the senses of to plait, um, in that sense]

Our findings also revealed other potential reasons for the mismatch between corpus data and intuitive data. It has already been widely observed that corpus data are obviously dependent on what was put into the corpus in the first place and do not represent 'language' as a whole (Leech 2007). The findings from this study highlight one problem with the BofE in particular. The corpus data showed a particularly high frequency for expressions related to sports journalism which were virtually absent from the informant data. For example, expressions such as 'thread the ball through' and its metonymic extension 'thread a shot through' may have artificially inflated senses 2 and 3 of thread as a verb (see Table 1). These senses represented 36.3% of the tokens in the BofE, but only 7.6% of the tokens in the intuitive data. Another problem with the BofE is that it appears not to reflect current usage. For example, 'eyebrow threading' was mentioned by 15% of the native speaker participants but only features once in our BofE data. Also, email discussion threads were under-represented in our BofE data in comparison with the intuitive data, particularly from the younger informants. The bias towards written texts in the CREA may likewise account for the frequency of occurrence of the conservative 'spin' sense of *hilar* in comparison to speakers' intuitions, and, in the case of both corpora, the under-representation of spoken language in general compounds the problem of representing new usages of words such as those studied here as they evolve, as these are likely to emerge in spoken discourse before being recorded in written form. The under-representation of spoken data also means that dialectal uses, such as one mentioned below by one informant, are not represented in the corpus, although they may very well form an important part of an individual's radial category of meanings for the verb *hilar*. In this case, we see that the speaker mentions a sense associated with an intransitive use of the verb:

Extract 4
Hilar. Em. Hombre, pues ¿qué estás tramando? ¿qué estás hilando? Eso lo escucho mucho por mi pueblo, lo de ¿qué estás hilando? ¿qué estás tramando?

['To spin. Um. Well, what are plotting? What are you spinning? I hear that a lot in my village, the thing about 'what are you spinning? What are you plotting?']

The fact that neither the corpus nor the intuitive data provide a perfect representation of 'general usage' indicates that there is a role for both approaches in studies of language. It also underscores the importance of individual differences and discourse communities in language study.

3.2.2. Corpus data and intuitions compared: Qualitative analysis

The main reason why the data produced by informants differs from that in the corpus is that it is susceptible to factors other than frequency. Ellis 2006 (a and b) lists a number of cognitive processes that are likely to be used by second language learners to turn L2 input into acquisition. These include: noticing, entrenchment, interference, over- and under-extension, probabilistic processing, contingency learning, learned (in)attention, salience, and perceptual learning. Many of these processes are likely to be employed by native speakers too as they develop their 'mental corpus'. For example, probabilistic processing refers to the remarkable sensitivity that learners have regarding the relative frequency with which certain forms are used in particular contexts in the input they receive, and their ability to match their output according to what they think might be appropriate. In other words, probabilistic processing can be seen as a kind of 'intuitive statistics' by which people assess the frequency of particular form meaning combinations and produce them in their own language. This process would explain the relative similarity between the corpus and the intuitive data. On the other hand, differences between the two are likely to be explained by some of Ellis's other processes. For example, perceptual learning may explain why certain alliterative expressions (such as 'wing its way') are more likely to be remembered than other, less alliterative phrases. In addition, certain form-meaning pairings are likely to be more easily attended to by certain individuals because they are more relevant to their everyday lives. Certain items are therefore likely to be more salient and people's interests and previous experience will lead them to attend to some features of the language and not others. The 'psychological prominence' or salience of stimuli, as Ellis points out, may be in part related to their physically measurable intensity, but essentially it refers to a subjective experience: 'Salience, as subjective experience, varies between individuals' (2006a: 16).

Indeed, considerable variation was found between the participants reflecting their age, gender and general interests. For example, message 'threads' on email discussion lists were only mentioned by participants below the age of thirty, as was the use of 'thread' as a form of facial hair removal. On the other hand, *The West Wing* (a popular television drama about the White House) was only mentioned by participants over the age of thirty. Military senses of 'wing' such as 'get your wings' were mainly mentioned by male participants. Among the Spanish informants, *hilar fino* was only mentioned by male participants. This shows how word knowledge is strongly related to background and suggests that the search for an idealised native speaker may not be that useful except for the contexts of dictionary compilation and very general English language teaching.

Gender differences were particularly noteworthy in the case of the Spanish informants. Among this group, the most striking difference was that all the informants who mentioned *hilar fino* were male; all the females interviewed mentioned either the sewing or spin senses of the verb. Furthermore, two male informants did not mention the concrete or basic senses of the verb at all, although they listed several figurative uses. Even when prompted ('And in the domestic sphere?'), neither produced the spin or sew senses of the verb, and one declared '*Pero eso es diferente*' ('But that's different'). The radial category knowledge of this group of informants varied quite substantially according to gender and the experiences that might accompany gender-specific roles. Senses associated with domestic chores such as sewing appeared to be less relevant to male informants, and none of the women remembered the idiom *hilar fino* ('to proceed or act with extreme caution and attention to detail'), whose use would be associated with public arenas, for example, negotiating agreement with sometimes hostile interlocutors.[3] Such differences in responses underline not only the dynamic usage-based knowledge of language, but also the value of drawing on the intuitions of different types of informants. While groups made up of informants with a very large number of shared characteristics (for example, first year psychology undergraduates) may provide interesting evidence in different experiments (with few troubling variables to skew experimental results), such testing may conceal important differences between individuals, which draws attention to the need to contrast evidence obtained from matched cohorts with groups of different characteristics.

3. It should be noted that all of the female informants were either engaged in full-time study or held a full-time job.

As has been mentioned, the Spanish informants' responses were not inhibited by the infinitive form of the prompt word, but they did show awareness of certain usage restrictions. This was particularly noticeable in informants' description of the idiom *hilar fino*, as all those who mentioned this sense also mentioned the positive evaluation implied by the phrase, echoing the uses found in the corpus:

Extract 5
Yo sólo utilizo la expresión en ese sentido de 'hilar fino' que es cuando alguien hace algo con mucha delicadeza, con mucha precaución y sobre todo con mucho tacto, lo utilizo para contextos del tipo por ejemplo social cuando alguien no quiere meter la pata y dice 'es que hay que hilar fino aquí' por ejemplo para no ofender a la gente. Es decir, hacer algo con mucho mucho detalle.

[I only use the expression in the sense of 'to proceed or act with extreme caution and attention to detail' which is when someone does something very cautiously and above all with great tact, I use it in contexts like social ones when someone doesn't want to put their foot in it and says 'you've got to go very carefully here' for example so as not to offend people. That is, to do something with great attention to detail].

Episodic memory appeared to play an important role, as did the presence of physical pain associated with the word thus giving an added dimension to the idea of 'perceptual learning', as we can see in Extract 6. Interestingly, this sense of 'threading', which was extremely salient for this participant, only appeared once in our Bank of English data:

Extract 6
Participant 19. I can also think of threading when I went to Jordan into a hairdresser's once and they threaded my face which was enormously painful.

Informants often mentioned very clearly recalled contexts of use or scenarios associated with senses of the verb. For example, one respondent described her knowledge of the 'spin' sense of *hilar* as arising from reading *Sleeping Beauty* in childhood, and recounted the whole of the scene when the princess pricks her finger on the spinning wheel and falls into her hundred-years' sleep. Another extract from the transcript illustrates how one speaker related the use of the verb with a whole scenario:

Extract 7
Y pues bueno hilar es en principio para mí, eh pues cuando cogen la lana que está en bruto, al principio, así pues el proceso de la formación en hilos. Y eso es hilar. Y bueno también está todo el proceso de enrollarlo todo en la bobina. O sea para mí es todo el proceso de cuando coges la lana y empiezas a hacer hilo.

[Well 'to spin' is mainly for me, um, when they get the wool, at the beginning, and so the process of making thread. And that's 'to spin'. And then there's the whole process of winding it round the bobbin. I mean for me it's the whole process of when you get the wool and begin to make thread]

Thus, frequency effects may account only partially for speakers' recall of senses associated with verbs such as these. As has been stressed (Ellis 2006b), frequency alone does not account for successful learning of any particular feature of language. In L2 learning, when taken together, frequency, salience, and contingency may all predict successful learning, but of these three factors, salience has the greatest predictive power (ibid., 173). This finding appears to be relevant to native speakers also. Particular contexts of use of these words appear to have proved particularly salient for speakers (associated with an especially memorable discourse event, experience or text), but a language corpus can give no insights into the relative salience of the usage events it records for the language user him/herself. Indeed, the fact that uses of the verbs *thread* and *hilar* in sports journalism were not mentioned by any of the informants suggests that language uses may leave little trace in memory. This may be due to the fact that readers may process such text in a content-oriented manner, and that the language forms used have low salience for language users in such a context.

In mentioning the different senses of *hilar*, informants used such expressions as 'básicamente' (basically), 'sentido figurado' (figurative sense) or 'por extensión' (by extension) in classifying the relative status of the senses they listed, showing that radial category knowledge of the type described by cognitive linguists is accessible to language users. This was particularly apparent in the responses provided by expert respondents, although not confined exclusively to them. However, in general, the responses of expert participants were considerably shorter than those mentioned by naïve informants, and showed a tendency to conflate senses in a way similar to that used by lexicographers in dictionary entries. The following extract illustrates a somewhat extreme example of this strategy:

Extract 8

Interviewer: Y para ti, ¿qué sentidos tiene la palabra, el verbo, hilar en español. ¿Cómo se emplea este verbo?

Participant: (Pause 5 secs) *Convertir la lana en un hilo con el que se pueda coser y hacer otros servicios.*

Interviewer: ¿Más significados? ¿Más usos?

Participant: Enlazar, unir, incluso coser.

[Interviewer: So for you what senses does the word, the verb 'to spin' have in Spanish? How can you use this verb?

Participant: Pause 5 secs. Turn wool into thread which you can sew with or do other things.

Interviewer: Any other senses? Other uses?

Participant: To link, to join, even to sew.]

Taylor (2010: p. 34) suggests:

> A person's I-language – the system of knowledge residing in her brain – is the product of her exposure to a set of [external] E-language events; her [internal] I-language is as it is because it was acquired through exposure to E-language. Conversely, the language that a speaker produces (that is, her contributions to the totality of utterances in the E-language) reflects her current I-language; E-language has the properties that it has in virtue of the I-language of its speakers. It seems only natural, therefore, that I-language should be studied from the perspective of the E-language that is both the product of I-language and the basis of its acquisition. The working hypothesis would be that I-language should be aligned as closely as possible with what is known about E-language.

However, the way that usage events may be mentally represented by different users may vary quite considerably depending, among other things, on age, expertise or even the cognitive style of the language user. The more abstract representation of senses apparent in the way these were described by expert informants contrasted with the concrete, context-dependent senses mentioned by naïve informants in the Spanish group. Naïve informants most often related uses of the verb to accompanying direct objects (e.g. *hilar ideas* or *hilar una conversación*) while expert informants would provide a synonym (eg *unir* or *enlazar*) without mentioning what could be figuratively 'joined' by this verb. Concrete senses of the verb were also expressed in gestures by two female bystanders: one older woman (aged

74) made downward circular movements alternating between both hands, rubbing the tips of thumb and forefingers together, in a movement that suggested feeding thread through the fingers; another made a wavy movement from right to left with the right hand, with the tip of thumb and forefinger pressed together, suggesting the movement of sewing with a needle.

Recency effects also appear to have shaped some participant responses, as we can see in Extract 9, where the participant's most recent experience appears to have partly shaped his response:

Extract 9
I've just come from teaching so 'wing it' comes to mind.

And in some cases, the responses given by the participants appeared to be restricted by the pressure of the testing conditions, as we can see in Extract 10:

Extract 10
Thread can mean about a million different things . . . Oh I had so many ideas when you first said it and now I've forgotten them all . . . under pressure . . . nothing's coming to mind

The fact that the ability to recall the senses appeared to be affected by factors such as recency, forgetting, pressured testing conditions, and varied linguistic expertise demonstrates the dynamic nature of category knowledge. Retrieval of senses seems to be closely related to contexts in which speakers have used the word or learnt its meaning. The data from participants is likely to be skewed for all of these reasons and it varies significantly according to individuals. One way of interpreting this is to see the intuitive data as being skewed and imperfect. Another way of seeing it is as containing important information about how language works in the minds of individuals.

3.3. Research Question 4: In what ways do language learners differ from native speakers in terms of their 'radial category' knowledge?

In order to answer this research question, we compared the responses given by the native speakers to those given by the non-native speakers. It will be remembered that the non-native speakers were all intermediate-advanced level, resident in the country where the target language was spoken.

We found that there was a large difference between the number of tokens provided by the native and non-native participants as we can see in Tables 1–5. These tables only contain figures for those senses that could be found in the corpus, although as we will see below some of the non-native speaker participants produced senses for *thread* and *wing* that were not present in the corpus. The difference between the two populations was found to be substantial but this finding needs to be treated with caution due to the small number of participants in the study.

In general, the non-native speakers tended to stick much more to the basic senses, and as we can see in Tables 1, 2, 4 and 5, the majority of senses for the English words were missing altogether from the non-native speaker data. If we compare these figures to the corresponding figures for native speakers, we can see here that the language produced by the non-native speakers was relatively impoverished. Indeed, far more senses for *thread* and *wing* were completely missing from the non-native speaker data than from the native speaker data. In fact the only non-basic sense of thread of which the non-native speakers were aware was the intransitive manner of movement sense ('he was threading through a minefield'). The non-native speakers had no knowledge whatsoever of *wing* as a verb. This finding can be partly attributed to the very small number of tokens produced by the non-native speakers. On the other hand the noun 'wing' is, in many ways, more basic and tangible than the verb 'to wing'. The fact that the learners were much more familiar with this sense is likely to be the result of frequency effects alongside a possible order of acquisition effect, in which more basic senses are acquired first. It has already been noted that non-native speakers tend to be wary of using figurative extensions of word meaning (Littlemore and Low 2006) but it is not clear whether this is because they do not know them or whether they are concerned about using them incorrectly. These initial findings indicate that the problem may simply be that they do not know them, possibly because they have failed to notice them. This indicates that sensitivity to frequency is not the main cognitive process in the acquisition of radial categories. Other processes, such as L1 interference, under-extension, and learned (in)attention may well go some way towards explaining the findings in these data.

On the other hand, the performance of the non-native speakers for the *noun* form of the two words was much closer to that of the native speakers. Both native and non-native speakers were aware of the internet sense of 'thread' which was relatively poorly represented in the Bank of English. Their knowledge of 'wing' as a noun was not so different from

that of the native speakers in terms of distribution of the senses (though in absolute terms it was relatively impoverished).

For both the noun and the verb forms, the non-native speakers occasionally produced senses that do not exist in English. Examples for 'thread' included:

There's a circus and a monkey and people can go on the thread (Participant 4)

The water is threading (Participant 10)

Thread is like a bunch of something (Participant 12)

Examples for 'wing' included:

I think when you go parachute it has wings (Participant 3)

Wing of a person ... if they're really confident you could say they have wings (Participant 4)

The wing of a kite (Participant 9)

When asked, the informants were unsure as to whether these senses were being transferred from their own language but in all cases they were convinced they had heard them before. What we appear to have here is a case of over-generalisation blended with transfer, and possibly entrenchment. Because of the way the radial categories are structured in their own language the non-native speakers appear to be convinced that the corresponding radial categories in English operate in the same way. This is similar to the schema effects that have been identified for native speakers in which people are convinced that certain elements are present in a story even when they are not, simply because the schema that they have developed for that type of story leads them to predict certain types of events. What we appear to have here is a sort of schema effect for radial categories.[4]

To sum up, non-native speakers of English produced significantly fewer senses than native speakers. However, compared with the corpus data, intuitive data for both native and non-native speakers were relatively impoverished. Non-native speakers had very limited knowledge of the

4. Here we refer to the more conventional use of the word schema and not the sense with which it is normally used within the discipline of cognitive linguistics. In other words, the word 'schema' here refers to the sets of expectations that people develop about certain types of language and genres, based on generalisations from their previous experiences and cultural conventions (see Cook 1997).

senses that lie towards the periphery of the categories compared with that of native speakers. Both groups produced more senses for nouns than for verbs and the senses produced by the non-native speakers were more similar to the native speakers for nouns than verbs, suggesting that they found the noun forms easier to learn than the verb forms.

The findings for the non-native speakers of Spanish were somewhat different from the non-native speakers of English. The group consisted of a total of nine informants, four of whom were undergraduate students from the same university in the United States, spending a year in Spain; the remaining five informants were all university lecturers who had been living in Spain for more than five years. The age of these informants and the length and type of exposure to Spanish account for the remarkable difference in their responses to the prompt. The younger informants did not recognise the verb *hilar* at all, and were unable to produce any meaning for it, although they reported that they had received between seven and twelve years' instruction in Spanish. They were taking part in an immersion programme and had been in Spain for six months at the time the interviews were carried out.

The older group of informants (1 male and 4 female) had different L1s: Italian, French (2), German and English. Unlike the findings for the non-native speakers with English as L2, there was no evidence for the role of transfer from the L1 among these informants' responses, which were similar to the native speaker informants': the core sense of 'sew' was mentioned by 4 respondents and the figurative extension *-hilar palabras-* was mentioned by all of them. This is perhaps unsurprising, given that academic discourse in general and the discourse of specialists in language and literature would be likely to make these informants familiar with these uses of *hilar*. More surprising was that the 'gender pattern' discerned among the native speakers was replicated by these highly proficient users of Spanish as a second language: the only respondent who mentioned *hilar fino* in this group was also male.

The number of informants in this group is too small to allow for any conclusions to be drawn, but the difference between the ways in which the two groups of informants have learnt Spanish appears to have an impact on the build-up of their radial category knowledge. Those experiencing instructed foreign language will rely to a great extent on the graded input they have been exposed to or taught (which might be somewhat impoverished). In these circumstances, learners may either have no category knowledge at all of particular words or show the kind of impoverished knowledge evident in the non-native speaker informants for *thread* and

wing. In these cases, impoverished or incomplete knowledge of the semantic potential of a target language form may lead to transfer effects, giving rise to over- and under-extensions of the senses possible, depending on what L1 form the user has matched it to and what the meaning potential of L1 form is. In contrast, those who have acquired the language in a naturalistic setting, with little or no formal instruction, will experience these language uses in similar ways to native speakers and there may well be fewer differences in their mental representation of the sense relations for verbs such as these.

4. Conclusions

To sum up, the main findings in this study were that English and Spanish differ in terms of radial categories, and corpus and intuitive data do not always match up, with the basic senses being more prevalent in the intuitive data than in the corpus data, while some figurative uses were present only in the corpus data. There was considerable variation across participants in terms of age, gender, and linguistic experience, and English-speaking participants seem to be more sensitive to the importance of parts of speech than the Spanish speakers. As for the differences between the native and non-native speakers, the non-native speaker informants produced far fewer senses, they focused even more heavily on the basic sense, and they were unaware of many senses.

The lack of correspondence between the corpus data and the intuitive data can be attributed to characteristics of both types of data. The intuitive data are susceptible to a range of phenomena, such as recency effects, the role of episodic memory, varied experiences, learned inattention, salience, and individual differences. The corpus data themselves are unlikely to be a perfect match for the intuitive data for several reasons. Firstly, the CREA and the BofE only contain approximately 10% oral language, so regional variation in the uses of these words is unlikely to be reflected. The written texts (though fairly well-balanced in type) will tend to reflect more conservative uses (topics may be the past/historical and literary texts may also talk of the past) which may explain the frequency of the 'spin' sense of *hilar*. In the BofE there is a strong weighting towards journalistic prose, particularly sports journalism, which may explain the skewed weighting of some senses of *thread*. More importantly, *thread*/*hilar* and *wing*/*aletear* are relatively low frequency items in both English and Spanish and one would expect them to be much more prevalent in some genres

than in others. Our study may well have yielded very different results if more frequent and more evenly dispersed words had been used. On the other hand, it is important not to overlook these less stable, less frequent items in language as they too can provide important information about how languages vary across contexts and genres. If language is viewed as a complex system (Larsen-Freeman and Cameron 2007) then it is important to give equal consideration to both its instable elements as well as its more stable features. If we simply attempt to 'iron out' the more 'difficult' and less evenly distributed elements of language then we risk creating an artificial picture of how languages actually behave. It would therefore be useful to replicate this study with more high frequency items and to investigate the differences between the findings produced by such a study and the findings produced by the study we have just described. Equally, it would be interesting to replicate this study with one that, instead of comparing speakers' intuitions about word senses with data contained in synchronic corpora such as these, compared these with data from L1 or L2 acquisition corpora. Future studies could usefully investigate intuitive data produced via more naturalistic tasks such as simple interviews or even language games.

Our tentative conclusion is thus that for low frequency items large synchronic corpora cannot give us much information about the relationship between senses in an individual speaker's mental grammar because experience in general and of language in particular is so variable that this will almost inevitably be reflected in individual speakers. On the one hand, corpora record usage events in a much wider range of circumstances than most people will encounter in their lifetime, though highly educated speakers may be familiar with many of the written genres used in compiling a corpus. On the other hand, an individual's knowledge of word senses is dynamic, constantly being up-dated and modified by experience in discourse events. The topics and registers that individuals are familiar with at any given point in their lives will affect not only the salience of certain words senses over others, but also the frequency with which they are encountered. Younger speakers are aware of different senses for words from older ones; experts know more about word meanings than non-experts; men and women do not always have the same sense relations. Thus, although we must be cautious about drawing conclusions based on a very informal method of eliciting data, it is hardly surprising to find that when we question a cross-section of a language-speaking population, as we have done here, we find that frequency of senses found in a corpus does not mirror the individual's notions of the meaning potential of such words.

What we have done here is simply to provide a snapshot of speakers' knowledge. These speakers' understanding of the many meanings of the words looked at will change in response to different experiences throughout their lives. We have a sufficient number of informants to give us some initial ideas, but the methodology now needs to be refined in order to tap into sense relations that people have but of which they may not be consciously aware. A refinement of the methodology might shed light on the question of why abstraction or conflation of word senses is favoured by older, expert speakers of Spanish but not by their counterparts when talking about English words. Cognitive style variables might play a role here, as might also transfer of training, in the sense that formal learning experiences, or the way individuals are trained to think about language, may affect the way sense relations are stored in the mental grammar. Thus, although speakers' intuitions may seem messy and provide unwelcome noise, this data may contribute to a more robust and cognitively-oriented description of the way that speakers organise their knowledge of word senses. Such descriptions would be especially welcome in contexts such as that of instructed second language acquisition.

References

Alejo, Rafael
2008 The acquisition of English phrasal verbs by L2 learners: a cognitive linguistic account. *Paper presented at the LAUD Symposium on Cognitive Approaches to Second/Foreign Language Processing: Theory and Pedagogy*. Landau, Germany, March 2008.

Brone, Geert, Kurt Feyaerts, and Tony Veale
2006 Cognitive linguistic approaches to humor. *International Journal of Humor Research* 19(3): 203–228.

Cameron, Lynne and Alice Deignan
2006 The emergence of metaphor in discourse. *Applied Linguistics* 27(4): 671–690.

Cook, Guy
1997 Schema. *English Language Teaching Journal* 51(1): 86.

Deignan, Alice
2005 *Metaphor and Corpus Linguistics*. Amsterdam: John Benjamins.

Dirven, Rene
1985 Metaphor as a basic means for extending the lexicon. In *The Ubiquity of Metaphor: Metaphor in Language and Thought*, Wolf Paprotté and René Dirven (eds.), 85–119. Amsterdam/Philadelphia: John Benjamins.

Divjak, Dagmar
2010 *Structuring the Lexicon: a Clustered Model for Near-Synonymy.* Berlin/New York: Mouton de Gruyter.

Earles, Julie L. and Alan W. Kersten
2000 Adult Age Differences in Memory for Verbs and Nouns. Aging, Neuropsychology, and Cognition. *Neuropsychology, Development and Cognition* 7(2): 130–139.

Ellis, Nick C.
2006a Language acquisition as rational contingency learning. *Applied Linguistics* 27(1): 1–24.

Ellis, Nick C.
2006b Selective attention and transfer phenomena in L2 acquisition: contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics* 27(2): 164–94.

Ellis, Nick C.
2008 Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 endstate. In *Handbook of Cognitive Linguistics and Second Language Acquisition*, Peter Robinson and Nick C. Ellis (eds), 372–405. Routledge: New York/London.

Evans, Vyvyan
2005 The meaning of *time*: Polysemy, the lexicon and conceptual structure. *Journal of Linguistics* 41: 33–75.

Giora, Rachel
2003 *On Our Mind. Salience, Context and Figurative Language*. New York: Oxford University Press.

Goatly, Andrew
1997 *The Language of Metaphors*. London: Routledge.

Goldberg, Adele E.
1995 *A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.

Goldberg, Adele E.
2006 *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Gries, Stefan Th.
2006 Corpus-based methods and cognitive semantics: The many meanings of *to run*. In *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*, StefanTh. Gries and Anatol Stefanowitsch (eds), 57–99. Berlin: Mouton de Gruyter.

Gries, Stefan Th.
2008 Corpus-based methods in analyses of Second Language Acquisition data. In *Handbook of Cognitive Linguistics and Second Language Acquisition*, Peter Robinson and Nick C. Ellis (eds), 406–431. New York: Routledge.

Gries, Stefan Th.
forthcoming Dispersions and adjusted frequencies in corpora: further explorations.
Gries, Stefan Th. and Dagmar Divjak
2009 Behavioral profiles: A corpus-based approach to cognitive semantic analysis. In *New Directions in Cognitive Linguistics*, Vyvyan Evans and Stephanie S. Pourcel (eds), 55–75. Amsterdam: John Benjamins.
Hanks, Patrick
1996 Contextual dependency and lexical sets. *International Journal of Corpus Linguistics* 1(1): 75–98.
Hanks, Patrick
2006 Metaphoricity is gradable. In *Corpus-Based Approaches to Metaphor and Metonymy*. Anatol Stefanowitsch and Stefan Th. Gries (eds), 17–35. Berlin: Mouton de Gruyter.
Kauschke, Christina and Prisca Stenneken
2008 Differences in noun and verb processing in lexical decision cannot be to word form and morphological complexity alone. *Journal of Psychological Research* 37: 443–452.
Langacker, Ronald W.
1987 Nouns and verbs. *Language* 63(1): 53–94.
Lakoff, George
1987 *Women, Fire, and Dangerous Things.* Chicago: University of Chicago Press.
Larsen-Freeman, Diane and Lynne Cameron
2007 *Complex Systems and Applied Linguistics.* Oxford: Oxford University Press.
Leech, Geoffrey
2007 New resources, or just better old ones? The Holy Grail of representativeness. In *Corpus Linguistics and the Web*, Marianne Hundt, Nadja Nesselhauf and Carolin Biewer (eds.), 133–149. Amsterdam: Rodopi.
Lindstromberg, Seth and Frank Boers
2005 From Movement to Metaphor with Manner-of-Movement Verbs. *Applied Linguistics* 26(2): 241–261.
Littlemore, Jeannette and Graham Low
2006 *Figurative Thinking and Foreign Language Learning.* Basingstoke: Palgrave MacMillan.
Littlemore, Jeannette and Fiona MacArthur
2007 A corpus-based analysis of *thread, hilar, wing* and *aletear. Metaphor and Discourse. A Special Edition of Culture, Language and Representation: Cultural Studies Journal of Universitat Jaume I* (5): 131–150.
Longe, Olivia, Billi Randall, Emmanuel A. Stamatakis and Lorraine K. Tyler
2007 Grammatical categories in the brain: The role of morphological structure. *Cerebral Cortex* 17: 1812–1820.

MacArthur, Fiona and Jeannette Littlemore
2008 A discovery approach using corpora in the foreign language classroom. In Frank Boers and Seth Lindstromberg. *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology*. Amsterdam: Mouton de Gruyter.

Moon, Rosamund
1998 Frequencies and forms of phrasal lexemes in English. In Anthony Cowie (Ed.) *Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press.

Perani, Daniela, Stefano F. Cappa, Tatiana Schnur, Marco Tettamanti, Simona Collina, Mario M. Rosa and Ferruccio Fazio
1999 The neural correlates of verb and noun processing – A PET study. *Brain* 122: 2337–2344.

Real Academia Española
2001 *Diccionario de la Real Academia Española. 22ª Edición.* Madrid: Real Academia Española.

Sandra, Dominiek and Sally Rice
1995 Network analyses of prepositional meaning: Mirroring whose mind –the linguist's or the language user's? *Cognitive Linguistics* 6(1): 89–130.

Sinclair, John
1991 *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, John and Ronald Carter (eds)
2004 *Trust the Text: Language, Corpus and Discourse.* Oxford: Routledge.

Spenney, Maria J. and William O. Haynes
1989 Semantic and phonological performance in adults learning novel object and action words. *Journal of Psycholinguistic Research* 18: 341–352.

Steen, Gerard, Ewa Biernacka, Lettie Dorst, Anna Kaal, Irene López-Rodríguez, and Tryntje Pasma
2010 Pragglejaz in practice: Finding metaphorically used words in natural discourse. In Graham Low, Zazie Todd, Alice Deignan and Lynne Cameron (eds), *Researching and Applying Metaphor in the Real World*, 165–184. Amsterdam: John Benjamins.

Taylor, John
2003 *Linguistic Categorization*. Oxford: Oxford University Press.

Taylor, John
2010 Language in the mind. In Sabine De Knop, Frank Boers and Teun De Rycker, (eds.) *Fostering Language Teaching Efficiency through Cognitive Linguistics*, 29–57. Berlin: Mouton de Gruyter.

Thornburg, Linda and Klaus-Uwe Panther
1997 *Speech act metonymies.* In: Wolf-Andreas Liebert, Gisela Redeker, and Linda Waugh (eds): *Discourse and Perspectives in Cognitive Linguistics*, 205–219. Amsterdam and Philadelphia: John Benjamins.

Tranel, Daniel, Coleman Martin, Hanna Damasio, Thomas J. Grabowski and Richard Hichwa
2005 Effects of noun-verb homonymy on the neural correlates of naming concrete entities and actions. *Brain and Language* 92: 288–299.

Tyler, Lorraine, Billi Randall and Emmanuel A. Stamatakis
2008 Corticol differentiation for nouns and verbs depends on grammatical markers. *Journal of Cognitive Neuroscience* 20(8): 1381–1389.

Xiao, Richard Z., Paul Rayson, and Tony McEnery
2009 *A Frequency Dictionary of Mandarin Chinese*. Oxford: Routledge.

Conversion and the lexicon: Comparing evidence from corpora and experimentation

Laura Teddiman

Abstract

In the present study, we explore native speaker sensitivity to lexical categories in English by targeting words that are ambiguous with respect to lexical category outside of a sentence context. In particular, we focus on those words that can be used either as a noun or as a verb, using corpora to determine word frequencies. Each time a word occurs in a corpus, it is used within a context that supports a particular interpretation, and this was reflected in the decisions each participant made. Participants completed an online category decision task, in which they were asked to decide whether a given item was a noun or a verb. Results showed that ambiguous words were more likely to be categorized as nouns when they occurred more frequently as a noun in context. This effect was not itself categorical, but instead was modulated by the relative frequencies of nominal and verbal occurrence. An offline task where words were rated on how noun-like they were generated similar results. Here, corpora can be taken to represent a measure of summed linguistic experience over time. The corpus, or experience, presents the contexts by which the lexical category of an item can be determined, and results from this experiment suggest that speakers are sensitive to this general information in the absence of a given context. That this information is ably used in a time-constrained task speaks to the power of corpora as resources for linguistic investigation.

1. Introduction

Lexical conversion is a common and productive word-formation process in English that allows a single word to be associated with more than one lexical category without the use of overt morphological marking (Plag 1999: 219; Plag 2003: 107). For example, the word *work* can be used as both a noun, as in *I have a lot of* ***work***$_{N}$ *to do*, and as a verb, as in I ***work***$_{V}$ *at the lab*. In these two utterances, it is the sentence context that helps us to determine how *work* is being used, that is, whether *work* is a noun or a

verb. In both cases, the meaning of *work* and the role it plays can be easily interpreted through the use of the overall context and its identity as a noun or a verb becomes clear. However, the way in which such words are interpreted in isolation is not. In the present study, we explore native speaker sensitivity to lexical categories in English and the role of experience in categorization.

1.1. Accounts of Conversion

There are three main proposals that seek to describe how categorically ambiguous words (e.g., *work*) are stored in the mental lexicon. The first is *lexical underspecification*, in which the root of the word is not specified in the lexicon at all. Instead, the lexical category of the whole word is realized only when the root is placed within a supporting syntactic environment (e.g., Barner and Bale 2002; Halle and Marantz 1994; reference to *functional shift*: Farrell 2001). That is, the mental representation, or lexical memory, of *work* in the mental lexicon does not specify whether the item is a noun or a verb.

Underspecification is contrasted with proposals that rely on storage. Storage-based proposals hold that there is at least one entry for each word in the mental lexicon for which the lexical category of the root is specified (e.g., Don 2004). The mental representation of a root includes syntactic information (e.g., lexical category) in addition to information about word meaning. In these theories, the lexical entry for each word in the mental lexicon contains information about how the word should be used.

A third, related theory, calls for the use of "zero-morphemes," unrealized morphemes that are used to derive words of one lexical category to another through the process of zero-derivation (e.g., Harley 2003; Lee 2009).

1.2. Lexical ambiguity in psycholinguistic research

While a productive area of study, psycholinguistic research into lexical ambiguity has tended to focus on ambiguities arising from multiple word senses rather than explicitly addressing conversion processes. In single word processing, the majority of studies have made use of lexical decision tasks. In a lexical decision experiment, the only question asked of a participant is whether an item they see (or hear) is an existing word or not. This task does not require participants to choose or recognize a particular meaning that must then be integrated into a surrounding context. Similarly, it does not require that participants distinguish between categorical uses of

the presented items. Many lexical decision experiments have reported an ambiguity advantage in reaction times, where words with multiple meanings are responded to by participants more quickly than those with a single meaning, at least when no further integration of lexical meaning into a surrounding context is required (Jastrzembski 1981; see also Rodd, Gaskell and Marslen-Wilson 2002). This is perhaps a consequence of an observation made by Onifer and Swinney (1981), who presented evidence that every meaning of an ambiguous word is activated in the earliest stage of processing for approximately 200 milliseconds, even when a disambiguating context is provided. In terms of simple lexical decision, this means that many meanings are available to determine whether or not an item is a word.

In studies of sentence reading, when it is necessary to choose one meaning of an ambiguous word over another, longer latencies are often observed for ambiguous items. These latencies can further be influenced by the frequency with which each meaning is expressed (Simpson and Burgess 1985). Duffy, Morris and Rayner (1988) conducted an eye-tracking study where ambiguous words were preceded by a neutral context and by a disambiguating context. The ambiguous nouns used in this study had two different meanings, and these meanings were either balanced for frequency or were non-balanced (one meaning was dominantly expressed). Results showed that when preceded by a neutral context, i.e., one that did not disambiguate the word meaning, the relative frequency of the two meanings influenced processing. When the word meanings occurred with similar frequency, participant gaze durations were increased in comparison to words with a dominant meaning. This was interpreted to mean that frequency influences the speed with which the items were accessed. In the case of non-balanced meaning pairs, the dominant meaning was accessed first and sentence integration has already begun before activation of the second meaning. Based on their results, Duffy et al. (1988) propose a model in which all meanings of ambiguous words are activated (as in Onifer and Swinney 1981), but where more frequent meanings are activated before less frequent meanings.

More recent studies have rigorously examined differences between homophony (multiple unrelated meanings) and polesemy (multiple related meanings/senses), finding that polysemous words show a processing advantage compared to frequency-matched unambiguous controls, while homophonous words do not (e.g., Klepousniotou 2002; Klepousniotou and Baum 2007). Klepousniotou (2002) suggests that the means by which we access homophonous meanings and polysemous senses differs. It is suggested

that different homophonous meanings of a single phonological word are stored as separate entries, while polysemous meanings are derived online from a single basic sense (for further discussion of gradient senses, see Brown 2008). Faster reaction times to polysemous words, when compared to homophonous words, arose from a lack of competition between presented target items. This hypothesis is compatible with earlier eye-tracking evidence that showed longer fixations on homophonous words than on polysemous words when disambiguating sentence information followed rather than preceded the target word (Frazier and Rayner 1990). The relatedness of polysemous senses was suggested to be sufficient for meaning to be extracted and for sentence processing to proceed, without deciding on a specific word meaning before continuing.

With respect to the current study, words that undergo conversion from noun to verb or verb to noun occupy an uncertain position, for while it can be argued that they undergo a morphological process, it is not clear how the meaning extension of the resulting word in combination with its new lexical category should be treated (Klepousniotou and Baum 2007). There is neurolinguistic evidence that nouns and verbs are processed differently, and indeed, that categorical information may be an important factor in the organization of lexical storage (e.g., Shapiro et al. 2005). The question of how categorically ambiguous words are processed then is a question of both sense ambiguity and structure.

2. The Present Study

In the present study, we ask how words – in particular, categorically ambiguous words – are understood or perceived in isolation. The first question asked through this research is whether speakers are able to determine the lexical category of words without a supporting context. While the items under scrutiny in this study are categorically ambiguous, results from the unambiguous items are also of interest. If speakers are unable to determine the lexical category of unambiguous items, then this would strongly support underspecification in the lexicon, and speakers would not be expected to be able to determine the lexical category for ambiguous items.

The second question asked is, if speakers are able to determine the lexical category of a given item, is this behaviour in any way affected by the frequency with which an item occurs as a noun or a verb? If speakers are sensitive to frequencies of occurrence within one category versus another,

then we expect variation in responses related to relative frequencies of each item as a noun and verb.

2.1. Category Decision Task

The first experiment focused on online participant responses to categorically ambiguous and unambiguous words in isolation. As in a standard lexical decision experiment, participants were shown single words on a computer screen and asked to make a decision about that word. In the category decision task, participants were asked to decide whether each item was a noun or a verb. Responses to both unambiguous and ambiguous items were of interest, although the target group was made up of categorically ambiguous words.

Participants Thirty-two native speakers of Canadian English took part in this study. Participants were drawn primarily from the University of Alberta undergraduate student population and were compensated for their time.

Apparatus The category decision task was scripted using PsyScope 1.2 and was presented to participants on an Apple Macintosh computer.

Materials Stimuli were composed of 45 unambiguous nouns (e.g., *bird*), 45 unambiguous verbs (e.g., *earn*), and 35 categorically ambiguous words. Half of the categorically ambiguous words occurred more frequently as nouns (e.g., *work*), and half occurred more frequently as verbs (e.g., *walk*). Raw frequency differences varied from 0 (the word was equally likely to be a noun or a verb) to approximately 5000 (biased towards either a noun or a verb). Three words fell outside of this range (verb: *need*, nouns: *time*, *school*). See Appendix 1 for a list of all tested words. An ambiguous word was considered to have a preferential reading as a noun if it occurred more frequently as a noun than as a verb. All frequency counts were extracted from the CELEX Lexical Database (Baayen, Piepenbrock & Gulikers, 1995). For the purposes of analysis, "errors" in categorization refer to responses that do not match the frequency-derived preferential reading. Figure 1 presents a schematic representation of the stimuli in this experiment. There were four types of words represented: unambiguous nouns (bird) and unambiguous verbs (earn), indicated by solid lines, and ambiguous words weighted towards a nominal interpretation (work) or verbal interpretation (walk), indicated by dashed lines.

Procedure Participants were seated in open cubicles before a computer screen. Stimuli were presented one word at a time on the computer screen.

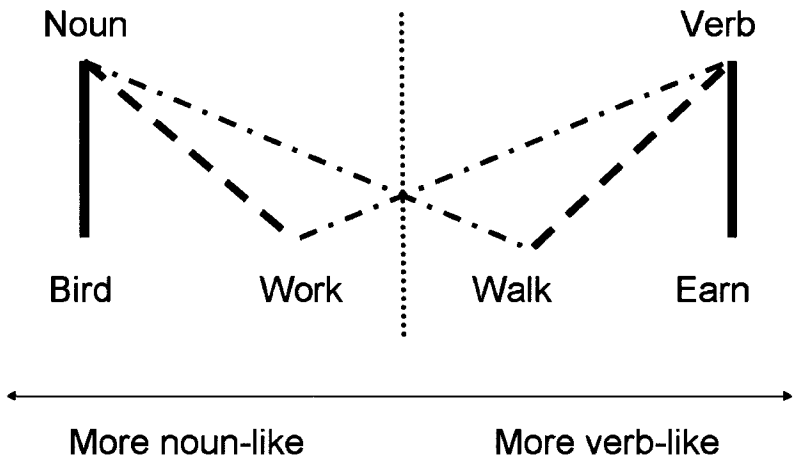


Figure 1. Schematic representation of stimuli used in Experiments 1 and 2

Participants were asked to decide whether each viewed item was a noun or a verb, and were encouraged to do so as quickly as possible while remaining accurate. To indicate their choice, participants pressed one of two keys on the keyboard. The "z" key, marked with a red sticker, represented a choice of "noun." The "/" key, marked with a green sticker, indicated "verb." Reaction times and error rates were recorded. On average, the experiment took approximately five minutes.

Results & Discussion The first question addressed in this experimental study was whether participants were able to correctly categorize unambiguous words as nouns or verbs. This question was important: if participants were unable to correctly categorize unambiguous words without a supporting sentence context, it would strongly support the underspecification hypothesis. Participants were able to categorize unambiguous items with 93% accuracy (Figure 2) even in the absence of a supporting context. This result indicates that for unambiguous items such as *earn* and *bird*, a sentence context is not required to determine the lexical category of these items. This suggests that the lexical category of the root in these cases is specified in the mental lexicon, or at the very least is readily available and quickly accessible in the absence of other information.

The second question addressed in this experiment was whether the relative frequency with which each item occurred as a noun compared to as a verb influenced participant choice during the category decision task. There was more variation in responses to categorically ambiguous words and so

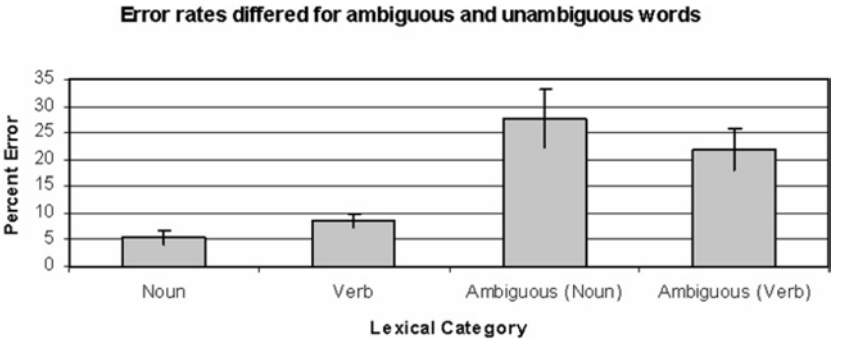


Figure 2. Participants were able to correctly indicate the lexical category of lexical items presented in isolation. Participants were less certain when categorizing ambiguous words, reflected in significantly higher error rates $F_2(3, 122) = 18.93$, $p < .0001$

there were significantly more errors in the categorization of the ambiguous target words when compared to categorically unambiguous items, as found by items using ANOVA, $F_2(3, 122) = 18.93$, $p < .0001$ and by subjects $F_1(1, 31) = 33.84$, $< .0001$. However, at approximately 75% accuracy, the categorization of ambiguous items remained above the level of chance. That is, 75% of the time, participants categorized ambiguous target items in accordance with the more frequently occurring lexical category for those items. If an item occurred more often as a verb than as a noun, then in general, participants were more likely to decide that it was a verb when required to make a decision. There was no significant difference in reaction times between ambiguous targets and unambiguous controls, $F_2(3, 122) = 2.3$, $p = .08$, and only the difference between unambiguous nouns and unambiguous verbs neared significance.

Categorically ambiguous words, however, are not in practice defined by a simple nominal or verbal preference. They vary in the degree to which they prefer one lexical category over another. Given that the preferred category for each ambiguous item was determined by using CELEX frequency counts, it was possible to quantify the categorical preferences of each word. To compare the relative frequencies of nominal and verbal occurrence, a simple measure was calculated by dividing the nominal frequency of each item by the total observed frequency (noun and verb) of that item. We have called this the *Celex Nominal Weight*. This essentially amounts to a percentage, and resulted in a number ranging from 0–1. A score of "1" indicated that an item identified as a noun only ever

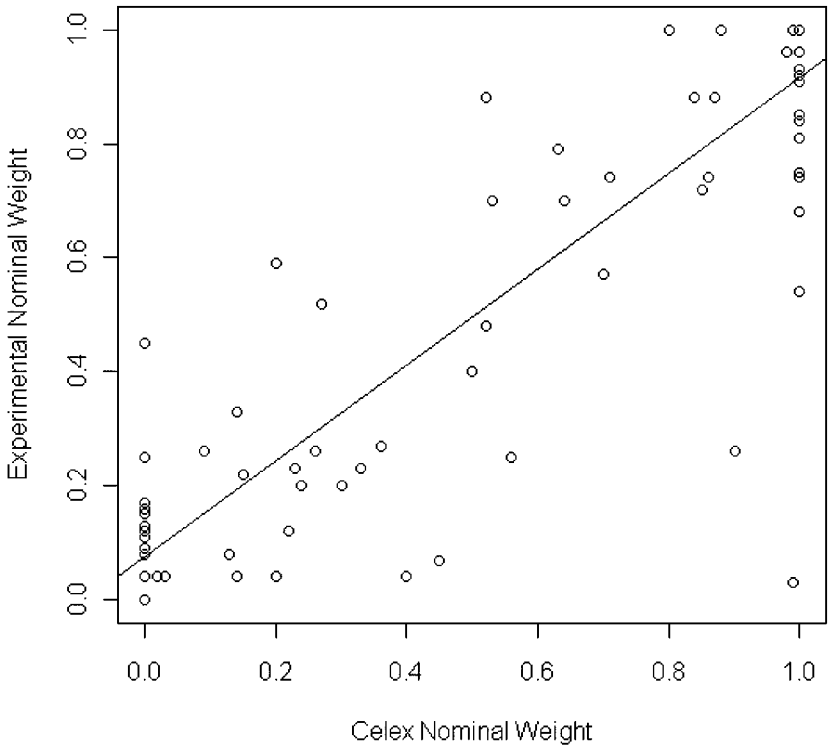


Figure 3. Relative frequencies of ambiguous words in CELEX are correlated with category choice, $r_s = .869$, $n = 125$, $p < .0001$. Each point represents a single word by its averaged means for Experimental and Celex Nominal Weights

occurred as a noun (a pure noun), while a score of "0" indicated a pure verb. A similar calculattion was performed to obtain the *Experimental Nominal Weight*, a number which measures how strongly an experimental item was perceived to be a noun during the online category decision task. To generate this number, the number of times a word was selected as a noun was divided by the number of times the word was observed during the course of entire experimental run. Looking at the ambiguous items alone using a Spearman correlation results in a correlation of $r_s = .573$, $n = 35$, $p < .001$, $R^2 = .328$. If we bring into the analysis the unambiguous items, then the correlation rises to $r_s = .869$, $n = 125$, $p < .0001$, $R^2 = .755$ (Figure 3). In order to further investigate the patterning of the observed experimental nominal weight, a linear regression was run on the data

using the *Celex Nominal Weight* and the total *Celex Word Frequency* as predictors for the *Experimental Nominal Weight.* The resulting model had an R^2 value of 0.867, and both predictors were found to be significant at $p < .001$. Participant responses recorded in the online task correlated with linguistic behaviour recorded in corpus data, and their behaviour was better predicted when accounting for word frequency in addition to categorical preferences.

Unambiguous stimuli had Celex Nominal weights closer to the extremes (0 = verb, 1 = noun), and typically, similar responses were recorded in the experimental data. More interesting were results from ambiguous items. Ambiguous items were more or less preferred as nouns in the frequency data extracted from CELEX, depending on the ratio of nouns to verbs for each item. This progression along a scale of "nominalness" is reflected in the experimental data. As words became more nominal as determined by the Celex Nominal Weight, more participants also decided that they were nouns in the category decision task, modulated by a whole word frequency effect. The decisions made by participants did not indicate that a higher nominal frequency for one category *always* resulted in the selection of that lexical category in the decision task. Rather, aggregating participant behaviour revealed gradient patterns similar to the behaviour of lexical items in the corpus data.

2.2. Offline Questionnaire

The second experiment was an offline questionnaire in which participants were asked to rate a series of words for how "noun-like" they were. After the binary decision required in the online task, this offline task was designed to give participants more time to consider their answers and more leeway in their final answers. With the use of the scale, we hypothesized that participants would rate categorically ambiguous items using the midpoints available, and that ratings would correspond to relative categorical frequencies.

Participants Twenty-three native speakers of Candian English participated in this experiment. Participants were primarily drawn from the University of Alberta undergraduate population and were compensated for their time.

Apparatus This task was a simple pencil and paper task. No special equipment was required.

Materials Critical stimuli were the same as those used in the category decision task, save that the lists were shorter than those presented in the experiment. There were two presentation lists.

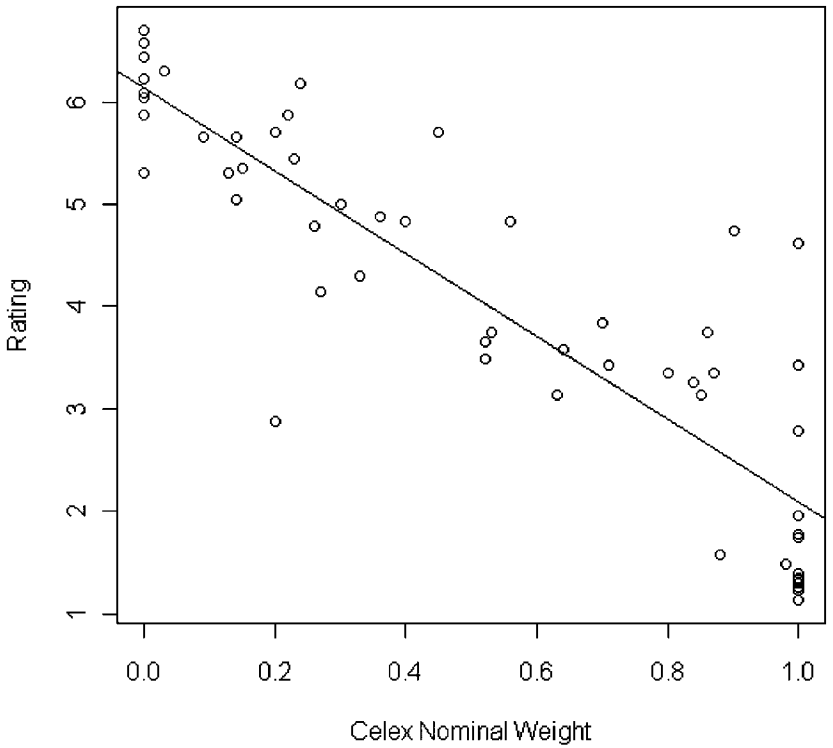


Figure 4. Relative frequencies of ambiguous words in CELEX are correlated with offline ratings, $r_s = -.877$, $n = 60$, $p < .0001$. Each point represents a single word by its averaged Rating and Celex Nominal Weight

Procedure Participants were given a list of experimental items consisting of unambiguous nouns, unambiguous verbs, and categorically ambiguous words, as in the category decision experiment. Instead of deciding whether each item was either a noun or a verb, they were instead asked to rate how 'noun-like' each word was using a scale of 1–7, where a rating of "1" indicated that the item was a noun (completely noun-like) and "7" indicated that the item was a verb (not at all noun-like). A rating of "4" indicated that an item could be interpreted as a noun or as a verb with equal ease. Participants were able to view the entire experimental list at the time of rating, and had the option of changing their answers. Most participants completed this task within ten minutes.

Results & Discussion Participants rated items largely in accordance with their relative categorical frequencies, as determined through CELEX. These

results mirrored those of the first experiment, as speakers rated nouns to be more noun-like when they occurred more frequently as nouns than verbs. The Celex Nominal Weight correlated negatively with the ratings reported in this experiment, as "7" was corresponded to "Verb-like" on the rating scale. Within just the ambiguous items, the Celex Nominal Weight correlated negatively with the ratings, $r_s = -.725$, $n = 35$, $p < .0001$, $R^2 = .526$. Using both ambiguous and unambiguous data points, this correlation grew to $r_s = -.877$, $n = 60$, $p < .0001$, $R^2 = .769$ (Figure 4).

The correlation observed here is stronger for ambiguous items than is observed in the online task and this difference could be related to the offline nature of the task. Participants were able to consider their answers and cross-reference other items on the list before completing this rating task. Unlike the timed category decision task, there was more opportunity for participants to make use of their explicit knowledge of English in determining how they treated each item. Another source of variation is the use of a scale. In the category decision task, participants were asked to make a binary choice, and aggregate participant choice reflected (or was reflected by) relative frequencies of noun versus verb occurrence in the CELEX Lexical Database. However, the scale used in the ratings task allowed participants to indicate how gradient they felt each word actually was, i.e., the choice was not binary. This allowed for more variation in responses.

2.3. General Discussion of Experimental Results

Thus far, two experiments investigating the processing of categorically ambiguous words have been described. In the category decision task, participants decided whether a given word was a noun or verb. In this binary decision, each participant only made one noun-or-verb decision per item, but when the aggregate responses were investigated, there was a correlation between the relative frequency of categorical occurrence and participant choice.

The results presented from these two experiments suggest that speakers are sensitive to the relative frequencies with which categorically ambiguous words occur. That is, speakers appear to be sensitive to whether or not a categorically ambiguous word occurs more frequently as a noun than as a verb, and this is reflected in their responses in both time-constrained online tasks and in offline tasks that allow for the use of metalinguistic knowledge.

Of note in these studies is that although context was not provided in either experiment, participants were still able to categorize experimental

items as nouns and verbs, although ambiguous words received more varied responses. This suggests that speakers may be sensitive to the probabilistic properties of words, and that probability may guide their responses. It does not provide support for lexical underspecification, where we would not have expected results in the category decision task to have diverged from chance. One might argue that the nature of the task forces participants to tap into contextual memory for the determination of lexical category, but that the words are themselves still underspecified for lexical category. If this were the case, we would expect there to be a marked difference in reaction times between the ambiguous and unambiguous words, as more time would be necessary to resolve the selection of category. This result was not observed. However, if we assume that the linguistic system is sensitive to probabilities of occurrence, with this information available at the lexical level, then participants would not require extra time to respond.

3. Generalizability

The experiments described herein used the CELEX Lexical Database as a source of frequency information. In order to test whether the results of these two studies were too closely married to the frequencies found in CELEX, another corpus was queried for frequency information. The goal was to test these results to see if they would generalize across another corpus that had not been used in the development of stimuli.

Davies' (2008) Corpus of Contemporary American English (COCA) is a freely available corpus of American English which contains more than 400 million words collected across a number of different genres. Although most participants in these experiments were speakers of Canadian English, as of this writing, there is no corpus of Canadian English readily available to academic researchers or to the public, and so COCA was used as a geographically representative corpus.

Relative frequencies recorded in COCA correlate with participant behaviour in the category decision task, just as those extracted from CELEX. When considering just the ambiguous stimuli, $r_s = .601$, $n = 35$, $p < .0005$, $R^2 = .361$. This correlation rises to $r = .897$, $n = 125$, $p < .000$, $R^2 = .805$ when examining the entire set of ambiguous and unambiguous items (Figure 5). Correlations between the ratings task and COCA frequencies were also significant ($r_s = -.738$, $n = 35$, $p < .0001$, $R^2 = .544$ for ambiguoius items; $r_s = -.848$, $n = 60$, $p < .01$, $R^2 = .72$, for all items).

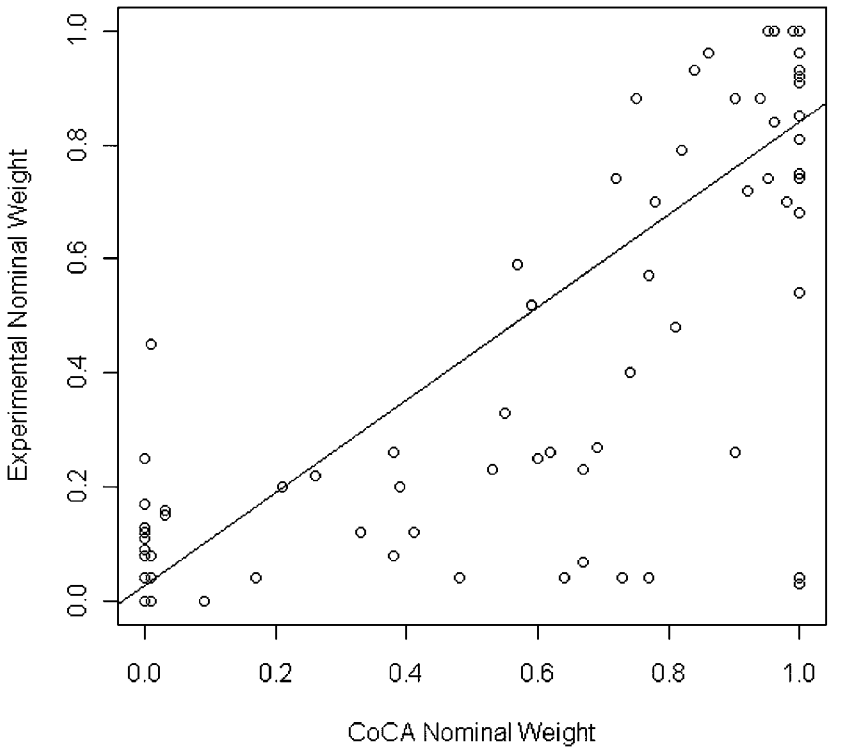


Figure 5. Relative frequencies of ambiguous words in COCA correlate with online experimental results, $r = .897$, $n = 125$, $p < .000$. Each point represents a single word by its averaged means for Experimental and Celex Nominal Weights

A further benefit to using COCA is that the corpus is broken down into several substantive genre-based subcorpora, and the search tool allows for investigation by genre. Categorical frequencies for each word used in Experiments 1 and 2 were collected for each major genre represented in COCA (Academic, Fiction, Magazine, Newspaper, Spoken). Among these, the experimental results best correlated with the Fiction (category decision task: for ambiguous items only, $r_s = .782$, $R^2 = .61$, for all items $r_s = .913$, $R^2 = .83$; rating task: for ambiguous items only $r_s = -.761$, $R^2 = .58$, for all items $r_s = -.854$ $R^2 = .73$) and Spoken subcorpora (category decision task: for ambiguous items only, $r_s = .745$, $R^2 = .55$, for all items $r_s = .899$, $R^2 = .81$; rating task: for ambiguous items only $r_s = -.727$, $R^2 = .53$, for all items $r_s = -.855$, $R^2 = .73$), while they differed the most from the Academic subcorpus (category decision task: for ambiguous items only,

$r_s = .66$, $R^2 = .44$, for all items $r_s = .879$, $R^2 = .77$; rating task: for ambiguous items only $r_s = -.627$, $R^2 = .39$, for all items $r_s = -.829$, $R^2 = .69$.)

With regards to the population tested, the study suffered a common drawback of psycholinguistic (and psychology) testing: reseach has been limited to the experimental population that is available and so the participants were, for the most part, undergraduate university students. They were generally well-educated speakers, although most students were beginning university and not finishing it.

This leaves the question of how well these results can generalize to different populations, and whether the familiarity with the concepts of "noun" and "verb" are responsible for the results observed here. There was no expectation that participants would be familiar with the classical definitions of nouns and verbs as they entered the experiment, and the instructions at the beginning of the experiment included basic definitions of both terms. The basic definition of noun was "person, place, or thing." The basic definition of verb was given as "action or state." Participants were largely drawn from an introductory linguistics class, but they had not yet completed the Syntax module of that class and so had not been exposed to linguistic analysis of that material. Furthermore, the first experiment did not allow participants to carefully consider their options, so an overt analysis of the materials is unlikely to have influenced those results. How other speakers from different educational backgrounds would respond in these types of experiments is, of course, an empirical question. One might hypothesize that differences in reaction times would emerge as participants less familiar with language analysis in general completed the task, and that response latencies would be longer overall. However, when there was no time pressure, we still observed correlations between frequency in a corpus and participant responses, so it is equally likely that offline responses would be similar to those observed here. With a corpus such as CoCA, which covers a wide range of topics and genres, it is likely that we have a reasonably good representation of the language environment of an average person, regardless of education. While less educated speakers may not have the same linguistic background as university students, the alternative here is to suggest that they do not share much, if any, of the same linguistic environment as the participants in this study, including those materials present on television, in magazines, and in newspapers. This is not to say, however, that CoCA and other large corpora will always be the best fit for each person individually. We might expect that specialized corpora frequencies would better capture the experiences of particular subgroups of the population.

4. General Discussion

Through these experiments, we have asked whether speakers of English are 1) able to categorize categorically ambiguous words as nouns or verbs and 2) whether the relative frequency with which words occur as nouns versus verbs influences their category choice. Speakers correctly identified the lexical category of unambiguous nouns and verbs, indicating that for these items, they have access to this knowledge outside of a supporting sentence context. Categorically ambiguous items were more difficult for speakers to categorize, but appear to still have a main categorical reading that is related to the relative frequencies of the noun and verb categories. Furthermore, variability rose as words became more ambiguous, as determined by the relative frequencies of the nominal and verbal forms of a word. These results suggest that frequency influences the storage of ambiguous items. Similar results were obtained through an offline rating task and were mainainted when another corpus was used. We interpret results from these experiments as evidence that speakers make use of their overall experience with a given lexical item in the absence of supporting context. Under this interpretation, words that undergo conversion are stored with probabilistic information specifying how likely a given word is to be encountered as a noun or a verb. This result is more closely related to the ratings task, where participants could indicate a gradient choice.

Although the results of these experiments show intriguing patterns with respect to categorization, there were no significant differences found in reaction times. In the online task, we might have expected elevated reaction times to ambiguous items because each of the target items was attested as both a noun and a verb. The lack of significance does not need to be surprising, however. The category decision task required participants to very quickly assign a lexical category (noun or verb) to the presented items. Participants were explicitly told that we were interested in their initial reactions to each word. Due to the speed of the task, there was not much time available for extensive metalinguistic analysis of each item. This task targeted "main," or dominant, categorical readings of each item. It did not target online derivation of items from one category to another. This differs from some context-based experiments, where reaction time differences and extended gaze durations are reported to semantically ambiguous words. Although most of studies of ambiguity have not addressed categorical ambiguity, some generalizations from their results can be useful for the interpretation of our data. Zemplini et al. (2007) investigated brain activity in response to semantically ambiguous words with a following disambiguation context. For target items, they used

semantically ambiguous words with dominant (more frequent) and subordinate (less frequent) meanings, embedded in sentences that supported one meaning over the other. Critically, the disambiguating context occurred after the target item. Their results suggested that the probability of a meaning occurring influenced whether or not it was selected before the context was revealed. If the incorrect meaning was selected, additional resources were required to update the meaning. Unlike Zemplini et al. (2007), in the present study there was no need for participants to 'update' the category or meaning of the target items, because there was no further disambiguating context. All that was required of participants was that they identify the dominant lexical category, although it was not phrased as such. Where participants did not select the dominant category, it is possible that they accessed the subordinate category first. Unfortunately, the extent to which this is true may vary according to the participant population and on the corpus selected for research.

While no corpus will match exactly the summed linguistic experience of any given speaker, we hypothesize that a corpus of sufficient size[1] can be used as a rough approximation thereof. The prevalence in a corpus of a given lexical item can be seen as a record of general language use. Reseachers do not have access to the total linguistic experience of each speaker, so large corpora are able to act as imperfect "stand-ins" for that background. Although no single speaker will have the exact language background represented in a corpus, the fact that corpora are drawn from existing language sources means that they should be able to mimic language exposure in a natural environment.

What is problematic about this approach to using corpus data in psycholinguist research is that methods of corpus construction can differ and word frequencies can be artificially inflated by just one or two source texts. Any individual source text is unlikely to be a perfect representation of language use. The major concern here is that the frequency with which an item is recorded in a corpus may be due to very few texts. A frequently

1. What constitutes a sufficient size is, of course, open to debate, and is highly dependent on the needs of the researcher. It would not be appropriate, for example, to use the Academic subgenre of CoCA to make predictions about language processing in children. For our purposes, the ideal corpus contains many genres, in order to capture the experiences of many and varied individuals, but is not so specialized in these subgenres that we target a small subsection of the population. A larger corpus is better in our circumstances because it is more likely that an average score will not be elevated by many occurrences in just one source text.

occurring word may occur in all subsections of a corpus equally, or it may occur very frequently in only a few texts. Gries (2008) discusses frequency and dispersion at length, and offers alternative measures to raw frequency. In the current study, the correlation between experimental evidence (category decision and offline ratings) and corpus-derived relative frequencies is stable across two large datasets (CELEX, COCA). The use of relative frequencies does not itself negate potential problems created by differences in dispersion patterns within a given corpus, but the replication in large datasets suggests that we can be reasonably confident that our results do not stem from a set of texts that are ultimately unrepresentative of the linguistic environment of our participants.

Our results suggest that educated native speakers of Canadian English are sensitive to the relative frequency with which words occur as members of different lexical categories. In general, categorically ambiguous words will be considered more noun-like if they occur more frequently as nouns, as measured through direct ratings tasks and a binary category decision task. If a large, well-balanced corpus is considered to be a proxy for the general linguistic experience of any given participant, then past experience with the linguist feature can be said to influence current behaviour. While this study targeted a small subset of words and asked fairly basic questions, the idea that corpora might be able to act as a stand-in for general linguistic experience is a powerful one. Corpora can identify the contexts within which given words are used. The frequency of that interpretation also informs how likely a given word is to be interpreted in that same way in the future. Speakers not only appear to be sensitive to this information, but also appear to use it, in this study when confronted with ambiguous items without any kind of disambiguating context. That this information is ably used in a time-constrained task speaks to the power of corpora as resources for linguistic investigation.

References

Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers
1995 The CELEX Lexical Database (Release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Barner, David & Alan Bale
2002 No nouns, no verbs: psycholinguistic arguments in favor of lexical underspecificaiton. *Lingua* 112: 771–791.

Brown, Susan Windisch
2008 Polysemy in the mental lexicon. *Colorado Research in Linguistics* 21: 1–12.
Davies, Mark
2008 The Corpus of Contemporary American English (COCA): 385 million words, 1990-present. Available online at http://www.americancorpus.org.
Don, Jan
2004 Categories in the lexicon. *Linguistics* 42(5): 931–956.
Duffy, Susan A., Robin K. Morris, & Keith Rayner
1988 Lexical Ambiguity and fixation times in reading. *Journal of Memory and Language* 27: 429–446.
Farrell, Patrick
2001 Functional shift as category underspecification. *English Language and Linguistics* 5(1): 109–130.
Frazier, Lyn & Keith Rayner
1990 Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language* 29: 181–200.
Gries, Stefan Th.
2008 Dispersion and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13(4): 403–437.
Harley, Heidi
2003 How do verbs get their names? Denominal verbs, Manner Incorporation, and the ontology of verb roots in English, Paper version of a 2001 talk at the Ben-Gurion University workshop on 'The syntax of aspect.'
Halle, Morris & Alec Marantz
1994 Some key features of Distributed Morphology. *MIT Working Papers in Linguistics* 21: 275–288.
Jastrzembski, James E.
1981 Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology* 13: 278–305.
Klepousniotou, Ekaterini
2002 The processing of lexical ambiguity: Homonymy and Polysemy in the Mental Lexicon. *Brain and Language* 81: 205–223.
Klepousniotou, Ekaterini & Shari R. Baum
2007 Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics* 20: 1–24.
Lee, Yongsung
2009 Conversion as zero affixation: Evidence from affix interaction. *Korean Journal of English Language and Linguistics* 9(1): 135–160.

Onifer, William, & David A. Swinney
1981 Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition* 9: 225–236.

Plag, Ingo
2003 *Word-formation in English*. Cambridge: Cambridge University Press.

Plag, Ingo
1999 *Morphological Productivity: Structural Constraints in English Derivation.* Berlin/New York: Mouton de Gruyter.

Rodd, Jennifer, Gareth Gaskell & William Marslen-Wilson
2002 Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language* 46: 245–266.

Shapiro, Kevin A., Felix M. Mottaghy, Niels O. Schiller, Thorsten D. Poeppel, Michael O. Flüß, Hans Werner Müller, Alfonso Caramazza, & Bernd J. Krause
2005 Dissociating neural correlates for nouns and verbs. *NeuroImage* 24: 1058–1067.

Simpson, Greg B. & Curt Burgess
1985 Activation and selection processes in the recognition of ambiguous words. *Journal of experimental psychology: Human perception & Performance* 11(1): 28–39.

Zempleni, Monika-Zita, Remco Renken, John C.J. Hoeks, Johannes M. Hoogduin, & Laurie A. Stowe
2007 Semantic ambiguity processing in sentence context: Evidence from event-related fMRI. *NeuroImage* 34: 1270–1279.

Appendix 1: word lists and raw word frequencies (CELEX)

Ambiguous nouns	Nominal Frequency	Verbal Frequency	Ambiguous verbs	Nominal Frequency	Verbal Frequency
block	976	425	curse	184	273
coach	615	103	dance	659	1177
fish	2927	557	drive	611	3907
glue	56	52	fall	856	5276
loop	149	60	goof	3	12
name	6012	1039	hope	1438	4080
phone	1118	643	kick	235	753
pipe	558	73	laugh	453	3058
plague	129	77	leap	141	549
school	9198	23	loot	65	65
screen	609	149	need	4178	9955
stain	228	214	nudge	26	95
stamp	332	298	paint	463	1285
state	5794	904	praise	243	302
time	35351	86	shrug	45	444
trade	3152	365	switch	391	780
work	12514	9925	walk	965	5552
			yawn	40	135

Unambiguous nouns: assistant (609), athlete (298), balcony (235), barn (226), beef (296), bell (745), bird (1841), boys (52), bulb (207), café (440), cairn (45), cards (135), chest (853), Congo (56), country (10015), desk (1633), dough (184), fate (615), fern (90), flight (1249), food (5317), gable (40), glee (65), gold (1575), heart (2937), injury (610), July (966), kelp (17), lamp (629), law (4021), lens (229), minx (3), moment (5965), pig (780), policy (4063), python (23), runner (556), satin (26), silk (474), street (5762), thug (86), trial (1185), trout (561), wisp (65), yard (1563)

Unambiguous verbs: accuse (648), adapt (464), adopt (1048), affirm (103), afford (1119), assist (362), avow (12), bask (69), begin (12254), bury (898), cease (659), clear (879), confess (456), consume (423), earn (973), eject (60), elicit (75), emerge (1433), endure (332), entitle (549), entrust (77), expire (77), fend (32), forbid (301), forget (3056), hear (9053), hover (214), hurl (224), learn (5524), lend (489), lessen (141), listen (3175), live (9371), lurk (154), perceive (391), quit (228), rend (29), scold (95), send (4822), sift (71), tend (2336), think (35874), wade (149), widen (241), wilt (54)

As lexical as it gets: The role of co-occurrence of antonyms in a visual lexical decision experiment*

Joost van de Weijer, Carita Paradis, Caroline Willners and Magnus Lindgren

Abstract

Antonyms are related in meaning but also frequently co-occur within each other's context. Taken together, these two characteristics determine the rating a pair of adjectives receives on a scale from very bad antonyms to very good antonyms. The antonyms that co-occur frequently (more often than chance would predict) and show a clear relatedness in meaning are the ones that receive the highest ratings in judgment experiments. The aim of the present study is to establish whether co-occurrence frequency independently can give rise to priming effects in a visual lexical decision task. For this purpose, we selected adjective pairs that are either antonyms or unrelated in meaning at either high or low co-occurrence levels in the British National Corpus. The results of the experiment show a priming effect for the antonyms but no effect for co-occurrence frequency. Frequently co-occurring pairs did not yield faster response times than infrequently co-occurring pairs, neither for the unrelated adjectives, nor for the antonyms. Our conclusion is that antonym canonicity is conceptual in nature, caused by the strength and the salience of the relation of opposition rather than the frequency of the lexical pairings.

1. Introduction

When asked about their opinion of how good a pair of lexical items are as antonyms, speakers judge *slow–fast* to be a pair of strongly antonymic

* This work was supported by a generous grant from the Swedish Research Council (www.vr.se) and forms part of a project called *Contrast in language, thought and memory*. We are grateful to Simone Löhndorf for her help with the experiments and to Roy Liddle for his comments on an earlier draft of this article.

adjectives, while *slow–rapid*, *slow–express* and *slow–blistering* are perceived as less good pairings and *fast–dull* as less good pairing than all the others. This raises the question *why* some pairs are considered better pairings than others and *why* they form entrenched pairings in memory and conventionalized couplings in text. There are most likely a number of converging reasons for goodness of antonym pairings (*antonym canonicity*), i.e. the extent to which antonyms are both semantically related and conventionalized as pairs in language. Such reasons may be the clarity of the meaning dimension, contextual versatility, symmetry, word frequency, frequency of co-occurrence, pair-wise acquisition, and stylistic co-occurrence preferences.

Co-occurrence has been shown to be an important factor in semantically oriented antonym experiments such as elicitation experiments and judgement experiments (Paradis et al. 2009 and Willners and Paradis 2010). It is well known that some words tend to collocate by virtue of the fact that they are related in meaning, as is the case with antonyms, or because the words are members of more or less fixed expressions, e.g. nominal constructions (*horse doctor*, *apple pie*, *great white shark*) or idiomatic expressions (*sick and tired*, *safe and sound*). Likewise, some words tend to co-occur in certain contexts or genres (*social* and *political*). In this study the focus is on the question of the importance of co-occurrence frequency for antonyms as well as for unrelated word pairs. The purpose is to find out whether co-occurrence frequency in itself can produce a priming effect from one word (an adjective) to its antonym (another adjective).

The way we investigate this question is schematized in Table 1. In a visual lexical decision task, we look at priming effects in antonym pairs and unrelated adjective pairs with varying degrees of co-occurrence frequency in the British National Corpus (the BNC). The example pairs in Table 1 illustrate the basic design of the study. We assume that co-occurrence frequency and adjective frequency are independent of one another to a certain extent. The adjectives *near* and *distant* are themselves more frequent than *vertical* and *horizontal*. Nevertheless, as a pair of antonyms, *horizontal* and *vertical*, co-occur far more often (390 times) than *distant* and *near* (11 times). Similarly, the adjectives *little* and *nice* co-occur 544 times, whereas *busy* and *plain* co-occur only twice.

Table 1. The experimental conditions

Frequency of co-occurrence	Antonyms	Unrelated word pairs
High	*horizontal-vertical*	*little-nice*
Low	*near-distant*	*busy-plain*

Section 2 reviews existing studies on antonym canonicity. In Section 3, our focus is narrowed down to co-occurrence frequency of antonyms. Section 4 states the research objective and the aims of this study and is used as a transition into the present study which is presented in Section 5. The results are discussed in Section 6.

2. Antonymy and canonicity

Antonym canonicity has recently been the focus of attention in a number of different investigations, using both textual and experimental techniques (e.g. Paradis and Willners 2007, Jones 2007, Jones et al. 2007, Murphy et al. 2009, Paradis et al. 2009, Willners and Paradis 2010). Antonym canonicity or goodness of opposability is the extent to which antonyms are both semantically related and conventionalized as pairs in language (Murphy 2003: 31). We argue that a high degree of canonicity means a high degree of lexico-semantic entrenchment in memory and conventionalization in text and discourse, and a low degree of canonicity means weak or no entrenchment and conventionalization of antonym pairs (Paradis et al. 2009). Antonym canonicity concerns lexico-semantic parings in language. The lexical aspect of canonicity concerns *which* words pairs are located *where* on an imagined scale from good to bad antonyms as measured in terms of participants' assessments and in terms of co-occurrence patterns in text, while the semantic side of the matter focuses on *why* some pairs might be considered better oppositions than others with reference to their conceptual set-up. In other words, while the lexical side of the coin concerns word forms, the semantic side of the matter concerns the characteristics of the meaning structure evoked by the word forms (Paradis 2010, Paradis and Willners 2011).

Paradis et al. (2009) and Willners and Paradis (2010) used English and Swedish adjectives respectively to measure which adjectives form part of strongly canonical antonymic relations and which adjectives have no strong candidate for this relationship. The method of data selection in both the English and the Swedish study was corpus-driven, sampled according to sentential co-occurrence frequencies. Two types of experiments were carried out using the pairings retrieved from the corpora – a judgement experiment and an elicitation experiment.[1] In the first part of the study (the judgement

1. Both studies yielded the same general results. In this article we will restrict ourselves to reporting on the English data from Paradis et al. (2009).

experiment), participants were asked to rate the goodness of oppositeness of adjective pairs on a scale from 1 to 11. In the second part (the elicitation experiment), another group of participants were given one adjective for which they were asked to provide the best possible opposite. The hypothesis under investigation was that there is a limited core of highly opposable couplings that are strongly entrenched as pairs in memory and conventionalized as pairs in text and discourse, while all other couplings form a scale from more to less strongly related pairings

On the basis of the data set used in Paradis et al. (2009), it was shown that the adjectives that were deemed to be excellent antonym pairs by the participants in the judgement experiment were also the ones, among the words searched for, that were the most frequently occurring in the BNC both as individual words and as co-occurring pairs. A set of canonical antonyms that differed significantly from the rest in the judgement experiment was identified. This result was subsequently confirmed by the results of the elicitation experiment, which showed that canonical antonyms elicited significantly fewer opposites than other adjective pairs. More generally, a strong correlation between adjective frequency and the number of antonyms suggested by the participants was found (Spearman rho $= -0.62$, $p < .01$) (Paradis 2010). In other words, more frequent adjectives tend to elicit fewer different antonyms than less frequent adjectives. The actual frequency for these adjectives is a sign of the fact that they may qualify a large range of nominal meaning structures and are useful in a large range of contexts. The fact that they are very frequent also calls for more research on that parameter, hence the topic of this investigation. It was also shown in the judgement experiment that sequential order does not play any role in the participants' assessments of goodness of antonymy.

Another factor that seems to be of importance for the best pairings, judging from the experimental results, is the salience of the dimension. The dimension of which the canonical antonyms are representatives is salient if it is easily identifiable. For instance, the SPEED dimension underlying *slow–fast* is easily identifiable, while the dimension underlying say *rare–abundant*, *calm–disturbed*, *lean–fat*, *narrow–open* are not. This has to do with the more specialized ontological applications of these adjectives to nominal meanings which concern different readings and sometimes also different meanings of these words and to certain very restricted styles and genres. This also means that polysemy and multiple readings within monosemy do not prevent a word from participating in a canonical relation with another word, e.g. *light–dark* and *light–heavy*, and *narrow–wide*

and *narrow–open*. Contextual versatility is a reflection of ontological versatility, i.e. the use potential of these antonyms applies in a wide range of ontological domains, and they are frequent in constructions and contrasting frames in text and discourse.

For the sake of the investigation, two approaches to antonymy were set up as contrasting positions by Paradis et al. (2009): the *lexical categorical* approach and the *cognitive prototype continuum* approach. The former approach considers antonymy to be a lexical relation and words are either lexical antonyms or not. Antonyms are pre-stored and get their meanings from the relation of which they are members. The model is context insensitive and static. Words either have antonyms or not. If they have antonyms they have one antonym. For instance, Miller and Fellbaum (1991: 210) state that *ponderous* is often used where *heavy* would also be felicitous, but unlike *heavy* it has no antonym. Similarly *heavy* and *weighty* have very similar meanings but different antonyms, *light* and *weightless* respectively. If antonymy was a conceptual relation, people would have accepted *weighty* and *light* or *heavy* and *weightless* as pairs of antonyms, which is not the case according to Miller and Fellbaum (1991). The conceptual opposition in their model between, say, *ponderous* and *light* is mediated by *heavy*. Conceptual opposition is thus an effect of lexical relations rather than its cause. However, the experiments carried out by Paradis et al. (2009) and Willners and Paradis (2010) paint a totally different picture. It is obvious, in particular from the elicitation experiment, that the participants have very different scenarios and different styles and genres in mind, when they offer antonyms to adjectives. The lexical categorical approach has no explanations for these patterns. Also, they predict a definite boundary between adjectives such as *heavy* that have antonyms and adjectives such as *ponderous* that have no antonyms on grounds that are not empirically supported. This predicts that we would obtain high scores which are consistent across native speakers for all adjectives that have antonyms and no responses for words with no antonyms, such as *ponderous*. In the lexical categorical model, antonymy as a category will be monolithic without any internal structure.

The cognitive prototype approach, on the other hand, takes antonymy to have a conceptual basis: antonymy is a construal rather than a pre-stored representation. It is dependent on general cognitive processes such as comparison and profiling and relies on a binary configuration of a segment of content (Paradis et al. 2009, Paradis and Willners submitted, see also Cruse and Croft 2004 for a construal approach to antonymy in the Cognitive Linguistics framework and Murphy 2003 for a context-sensitive pragmatic

approach). Adjectival meanings are evoked in conceptual combinations with nominal meanings. Conceptual structures are the cause of antonym couplings, not an effect, and salient contentful dimensions such as SPEED, LUMINOSITY, STRENGTH, SIZE, WIDTH, MERIT and THICKNESS form good breeding grounds for routinization of lexical pairings (Herrmann et al. 1986). This approach predicts a category with an inherent continuum structure with a small number of core members associated with particularly salient dimensions. The results of the investigation carried out by Paradis et al. (2009) and Willners and Paradis (2010) indicate that strongly canonical pairings have lexical correlates, while the vast majority of antonyms have only associatively weak partners in situations where speakers are invited to produce or evaluate antonyms without any contextual constraints. Given a specific context, antonym couplings are bound to be stronger and more consistent across speakers (Murphy and Andrew 1993). In the lexical categorical model different contexts do not affect the antonym, since the antonym of a word is not determined by context and sense, but is lexically driven. Finally, the prototype continuum model is consistent with categorization in general (Taylor 2003).

In line with the reasoning and the findings of Murphy and Andrew (1993) and Murphy (2002), the theoretical implication of our investigation is that antonymy is primarily a conceptual relation, based on general knowledge-intensive cognitive processes. However, the investigations also point to the fact that a select group of antonyms are particularly strongly associated in memory. They are deemed to be superb examples of antonyms by participants in judgement experiments and there is strong agreement across participants in elicitation performance about the best antonym of a given word from this select group. For instance, even though it is easy to produce possible antonyms of *bad* (*satisfactory*, *beneficial*, *fine*, *obedient*), all of the experiment participants offered *good* as the best antonym of *bad* in the above-mentioned elicitation experiment. Pairings for which the participants suggested many different antonyms in the elicitation experiment are more likely to be contextually limited, i.e. not strongly routinized as pairs in our minds, or very weakly conventionalized, more generally, due to extreme genre or register restrictions.

In spite of the fact that the test items in both the judgement and the elicitation experiments were presented out of context, the experimental types relate more to the semantic side of the pairings than the formal side. This is also the case in other priming experiments such as Becker (1980). In his visual word recognition experiment, Becker used two types of stimuli: antonyms, such as *smart–dumb*, *dry–wet* and category-name/

category-member pairs, such as *furniture–chair*, *dog–collie*. They were presented both as related, i.e. *smart–dumb*, and as unrelated, i.e. *smart–dry*, and the same design was used for the categorically related pairs. The cue-target materials were designed to produce a situation in which the participants were to predict consciously what a related target would be. The problem under investigation concerns the conditions under which a facilitating effect of an appropriate semantic context dominates and the conditions under which an interfering effect of an inappropriate context dominates. The experiment using antonyms produces a substantial facilitation effect and negligible interference, while the category-member relationship yielded only nominal facilitation but substantial interference. Like Paradis et al. (2009), Becker's investigation also shows that the order of the test items is of no importance. What is of importance, however, are the qualitative distinctions in terms of strength of relatedness. Both in the case of what Becker refers to as high-typicality antonyms (strongly canonical antonyms in our terminology) and high-typicality category-member pairings, relatively small interference effects but substantial facilitatory effects obtain. His conclusion is that the crucial factor is the type of stimuli. Facilitation dominance prevails for the condition that contains consistently strongly related test items, while interference dominance obtains when the test items are characterized by a wide range of semantic relationship strengths.

In sum, there are indications in the above-mentioned investigations that co-occurrence frequency is of importance for antonym canonicity. We take this as the springboard into the next section which concerns the current investigation of the role of frequency for strength of lexical affinity of antonym couplings.

3. Antonyms and frequency

There is widespread agreement in the literature that word frequency and semantic relatedness have facilitating effects in visual lexical decision performance (e.g. Becker 1979, 1980, Perea and Rosa 2002a, 2002b). Frequent words are recognized at higher speed than infrequent words, and targets preceded by related primes, e.g. *table–chair*, are known to be recognized faster than targets preceded unrelated primes, e.g. *table–moon*. Similarly, as already mentioned above, canonical antonyms have been found to prime each other more strongly than other opposites or words that are related through category-membership, such as *furniture–chair*, *dog–collie* (Becker 1980). It has also been shown in semantic priming experiments that low-frequency targets produce larger priming effects

than high-frequency targets (Becker 1979, Plaut and Booth 2000). In most of the literature, frequency refers to individual word frequencies rather than to co-occurrence frequency. There are a few exceptions to this in the priming literature, however. For instance, using different kinds of related words, such as synonyms and various types of category-membership relations, Spence and Owens (1990) demonstrate that co-occurrence frequency in text is significantly correlated with strength of association in priming experiments. In contrast to these findings, Estes and Jones (2009) show that co-occurrence frequency does not play a role in the explanation of integrative priming in expressions such as *lemon cake*, *horse doctor* and *plastic toy*.

The reasons for using a corpus-driven method for stimulus selection in Paradis et al. (2009) and Willners and Paradis (2010) and also in this study are twofold. One reason is to be able to establish the frequency levels for antonyms in naturally occurring, non-manipulated text and discourse. The other reason is to select test items for the experiments in a principled way using natural language since previous corpus studies have shown that textual evidence supports degrees of lexical canonicity. Charles and Miller (1989), Spence and Owens (1990), Justeson and Katz (1991, 1992), Fellbaum (1995) and Willners (2001) have established that members of pairs perceived to be canonical tend to co-occur at higher than chance rates and that such pairings co-occur significantly more often than other semantically possible pairings (Willners 2001).

The same assumptions are made in corpus-based treatments of antonym co-occurrence in text (Jones 2002, 2007, Jones et al. 2007, Muehleisen and Isono 2009, Murphy et al. 2009). These studies concern the aspect of frequently co-occurring antonyms serving various contrasting discourse functions in text in constructional frames such as 'neither X nor Y', 'X instead of Y', the difference between X and Y'. The results attest to the fact that antonym pairs which are perceived to be good opposites occur frequently in such discursive frames. Similar studies have been performed on data from the CHILDES database (Jones and Murphy 2005, Murphy and Jones 2008, Tribushinina (in press). For instance, using the American English component of the database, Murphy and Jones (2008) observe that children use antonyms at earlier ages than experimental studies have shown and they also use antonyms for mostly the same discursive purposes as adult speakers do. It has also been shown that lexical access and various levels of lexical priming may influence word choice and word prediction (Yap et al. 2009) and lexical priming may also be a triggering factor for speech errors (Söderpalm 1979, Gainotti et al. 1983, Varley 1991, Samson et al. 2007).

As was previously mentioned, co-occurrence has also been used in integrative priming experiments, i.e. testing the relations between two nouns, where the first noun is a modifier of the second noun and thereby designates a subcategory of the head noun in that they jointly name the category (Estes and Jones 2009). For instance, the concepts for *table* and *vase* may be integrated through a relation of location (*table vase*), through causation (*rope burn*), through composition (*copper pot*), through time (*winter holiday*), through function (*sketch pen*), through partonymy (*bear paw*), through topic (*cowboy film*) and through production (*wind power*). Estes and Jones (2009) examined McKoon and Ratcliff's (1992) argument that the frequency of co-occurrence in samples of written text is the best estimate of prime-target familiarity but found that co-occurrence in text did not explain integrative priming. Instead, Estes and Jones (2009) explain the mechanism of integrative priming using role assignment, e.g. location or causation, and propose that role typicality, relation plausibility and compound familiarity are crucial factors.

4. Research objective and aims

The results of the studies reviewed in the previous sections suggest a correlation between co-occurrence frequency and 'goodness of antonymy' of the form-meaning pairings. In line with these results, we assume that antonyms that co-occur often tend to be assessed as better pairs than antonyms that do not occur often. Even though more frequent adjectives generally co-occur more often than less frequent adjectives, we make the assumption that both co-occurrence frequency and adjective frequency may have independent facilitating effects on word recognition.

Our aim in this study is therefore to find priming effects that are due to co-occurrence frequency independently of individual word frequency. We are hereby able to evaluate the relative importance of co-occurrence frequency for the antonymic relations on the hypothesis that frequently co-occurring antonyms are more strongly conventionalized as pairings in text and discourse. The prediction that follows from the hypothesis is that frequency of co-occurrence of pairings speeds up word recognition. Indirectly, the outcome of such an experiment allows us to evaluate the two approaches to antonymy, the *lexical categorical* approach and the *cognitive prototype continuum* approach that were described in Section 2. The second aim is to establish whether co-occurrence frequency can cause a priming effect when two words are semantically unrelated. For instance,

the adjective *nice* will be recognized faster when it is preceded by *little*, since these two words often co-occur, whereas there is no facilitating effect for the word *plain* when it is preceded by *busy*, since these two adjectives are unrelated (see Table 1). The inclusion of semantically unrelated word pairs also allows us to examine possible interaction effects between co-occurrence frequency and semantic relatedness. At a more general level, co-occurrence frequency has rarely been used as a factor in priming experiments. Our results, therefore, complement existing findings on priming, and should therefore have implications for models of priming.

5. The experiment

This investigation combined corpus data with behavioural data. We carried out a visual lexical decision experiment for which we selected material, based on frequency information obtained from the BNC. In a visual lexical decision task, words and nonsense words are presented to participants on a computer screen. The participants' task is to press one of two buttons on a button box for a real word and the other button for a nonsense word. They are encouraged to respond quickly and accurately. Their reaction times are recorded, as well as their errors (incorrect button presses). Notable factors that influence reaction times are, among others, word frequency (high-frequency words are recognized faster than low-frequency words) and word length (short words are recognized faster than long words (cf. New et al. 2005)). We controlled for these factors in order to avoid interferences with potential co-occurrence frequency.

A total of 20 participants were tested: 17 women and 3 men, aged 19 to 40 years (average age was 24.75 years). Most of the participants were exchange students at the universities of Lund and Växjö in Sweden. They were recruited through the International Offices at the two universities. All of them had English as their native language from Australia (1), Canada (1), South Africa (1), Great Britain (7) and the United States of America (10). The reason for choosing the BNC as a source for data retrieval was a matter of the size of the corpus rather than the fact that the data are British English. We expect educated native speakers from any English-speaking part of the world to be equally familiar with the test items we used for the experiment, which are all frequent, 'common core' adjectives in all of the above dialects (see Appendix A). We obtained the participants' written consent to use their input for our investigation. At the end of the experiment, they received a lottery scratch card as compensation for their time and effort.

Table 2. Most frequent adjective pairs in the BNC

new	old	3,946
economic	social	3,765
black	white	3,498
new	other	3,493
economic	political	2,816
political	social	2,787
other	social	2,503
large	small	2,453
different	other	2,377
local	other	2,259

Materials

The adjective pairs that we used as primes and targets in the current experiment were selected on the basis of frequency counts in the BNC. The adjectives that were considered to be suitable items were all relatively common, with a frequency of at least 1,000 occurrences in the BNC. We compiled a list of all the adjective pairs that co-occurred within a sentence. The list consisted of 422,499 pairs with token frequencies ranging from 1 to 3,946 (a total token frequency of 4,454,280). Table 2 shows the ten most frequent adjective pairs. From the list, we manually identified 233 antonym pairs of which we chose 60 for the experiment, and decided which of the two adjectives would serve as the prime and which would be the target. Subsequently, we chose another 60 unrelated adjective pairs from the same list and matched these with the antonym pairs so that the targets had approximately equal length and frequency (see Table 3). No adjective occurred more than once in the experiment. The adjective pairs are shown in Figure 1. Note that the antonym pairs appear to have slightly higher co-occurrence frequencies than the unrelated. We will return to this issue in the analysis.

To the 240 adjectives, we added 287 phonotactically appropriate but non-existing English words, e.g. *goast*, *solt*, *voddle* and *foose*, yielding a total of 527 items. The item pairs were ordered randomly and mixed with the nonsense words in such a way that there were always one or two nonsense words between a target and a prime, and either zero, one, or two nonsense words between a prime and a target. The gap between primes and targets was varied randomly in order to make the appearance of the target word unpredictable for the participants.

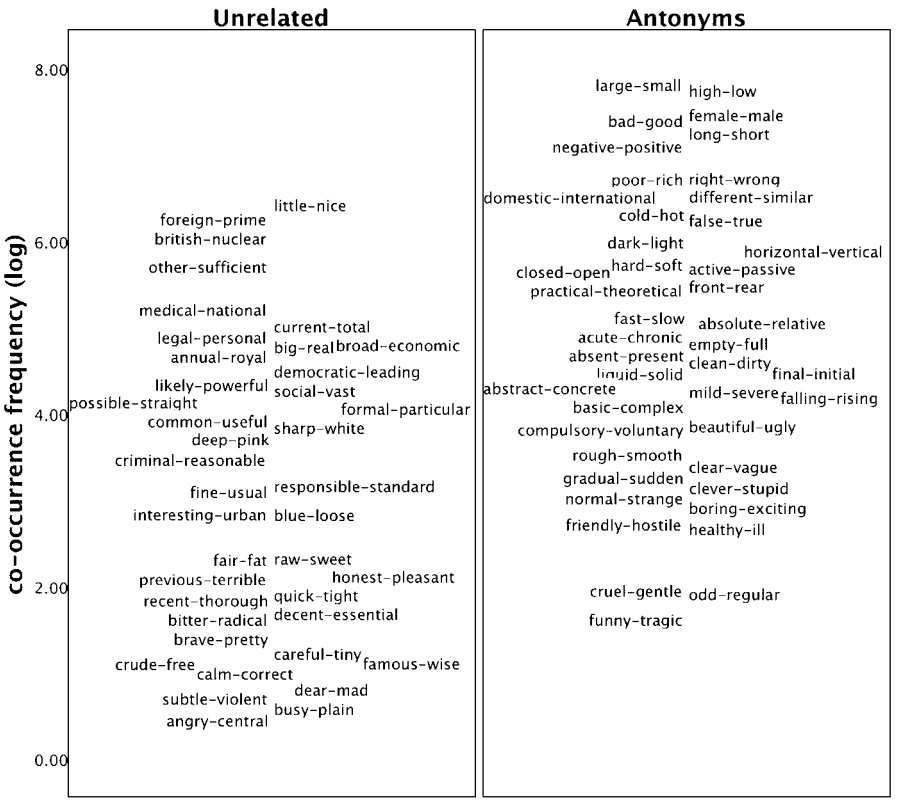


Figure 1. Experimental item pairs. The vertical positions of the adjective pairs correspond to their approximate frequency of co-occurrence

Experimental procedure

The presentation of the stimuli and the collection of the participant responses were controlled using E-prime software (Schneider, Eschman and Zuccolotto 2001). An experimental trial consisted of a time interval of 1500 ms during which the word was presented on the computer screen followed by a 500 ms empty screen. The participants were instructed to press either one of two buttons on a button box for a real word or a nonsense word. Only responses during the time that the item was on the screen were recorded. Responses given outside that interval were counted as errors.

The experiment was divided into five blocks. The first block consisted of 10 practice items (not included in the 527 experimental items) to make sure that the participants understood and followed the instructions correctly.

After that, the experimental items were presented in four approximately equally long blocks, giving the participants the opportunity to relax for a short while, and then continue with the next block when they felt ready to do so. The participants were tested in a silent room.

Analysis

In the analysis, we focused exclusively on the reaction times to the target adjectives. We were mainly interested in two factors: co-occurrence frequency (a continuous variable) and relatedness (antonyms or unrelated). In addition to these two main factors, we also looked at potential confounding factors: the individual frequency of the target, the length of the target, (the number of letters), the trial number (the moment at which a particular item appeared in the experiment) and the gap size between prime and target (0, 1 or 2 nonsense words). We had two concerns that were importance for our analysis of the data. One was the effect of possible interference of the confounding factors and the main predictors. Another concern was that co-occurrence frequency was slightly unbalanced across the experimental items, with the antonyms having a higher average co-occurrence frequency than the unrelated adjectives (see Table 3).

Table 3. Descriptive statistics for the target words. Frequencies are log-transformed values

	Antonyms	Unrelated
co-occurrence frequency	4.97	3.43
target frequency	8.77	8.99
target length (letters)	5.97	6.02

In order to overcome these concerns, we disentangled the effect of the main predictors and the effects of the confounding factors and analyzed the results by fitting a multilevel model to the data. Due to its flexibility, the multilevel approach had several advantages. First, the effects of the main predictors are tested while controlling for the effects of the confounding predictors. Second, unbalanced data (unless severely unbalanced) are unproblematic for the analysis (Singer and Willett 2003). Finally, frequency and word length are continuous variables. This would have been problematic for an analysis of variance but was not for the multilevel approach where predictors may be categorical or continuous.

In the model, reaction time was the dependent variable, which is predicted from the main factors, the confounding factors, and two random factors (i.e., participant and stimulus word). We measured the interaction of co-occurrence frequency and relatedness but also looked at their main effects. The predictions were straightforward: we expected a (semantic) priming effect for antonyms but not for unrelated adjective pairs.

Results

The first row in Table 4 shows the percentages of correct responses for the targets. Overall, the participants made remarkably few errors. The overall percentages of correct responses approached 99% both for the antonymic and the unrelated targets. Incorrect responses were excluded from further analysis. The second row of the table shows the average reaction times for the antonymic and the unrelated targets. In line with what we expected, responses to antonymic targets were faster than those for the unrelated targets by nearly 20 ms.

Table 4. Percentages correct and average Reaction Times

	Antonyms	Unrelated
Correct (%)	98.6	98.5
RT (ms)	572	591

The relation between co-occurrence frequency and reaction times is shown in Figure 2. For this figure and for the subsequent analysis, the reaction times and co-occurrence frequency were log-transformed in order to reduce skewness and the effect of outliers (Baayen 2008: 71). The slopes of the regression lines in the figure suggest a very slight negative relationship for the antonym pairs, and a very slight positive relationship for the unrelated adjective pairs.

The analysis started with an initial model containing all predictors: trial, item frequency, item length, lag size (0, 1 or 2 words), and the joint effects (i.e., the interaction plus the main effects) for relatedness and co-occurrence frequency. The results are given in Table 5. This table gives the estimates of the effects, together with their standard errors. The sign of a coefficient (plus or minus) indicates the direction of the relationship between the predictor and the dependent variable (positive or negative). The last column shows the accompanying *t*-values, indicating whether the

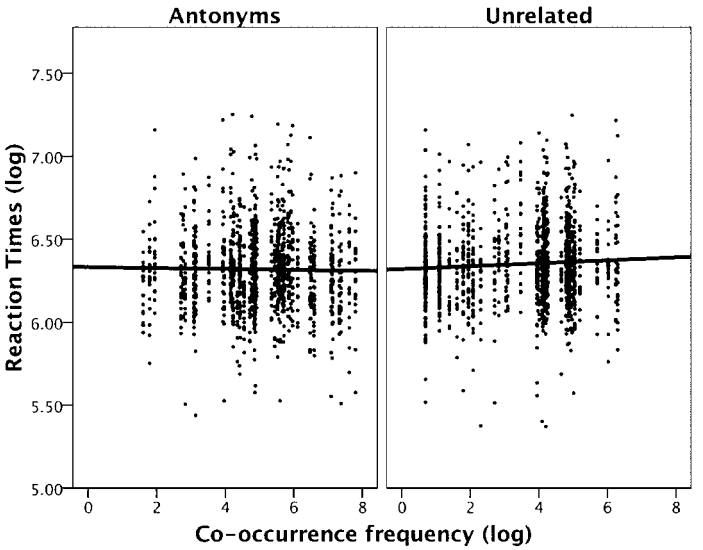


Figure 2. Relationship between co-occurrence frequency and reaction times

Table 5. Initial model

	coefficient estimate	standard error	*t*
trial	−0.0002	0.0001	−2.45*
relatedness	0.0058	0.0376	0.15
co-occurrence frequency (log)	0.0030	0.0063	0.48
co-occurrence freq × relatedness	0.0082	0.0082	1.01
target frequency (log)	−0.0208	0.0079	−2.64*
lag size	−0.0016	0.0079	−0.20
target length	0.0114	0.0033	3.44*

effect of the predictor is significant or not. Since the number of observations is rather large (over 2,000), the *t*-values may be interpreted as *z*-scores, and thus absolute values larger than 1.96 may be considered significant with $p < .05$ (cf. Baayen 2008: 248).

The *t*-values given in the rightmost column suggest significant effects of trial, target length, and target frequency, but not of relatedness, lag size, co-occurrence frequency nor of their interaction. In the next step, we

excluded the interaction term and tested the individual effects of relatedness and co-occurrence frequency by fitting two new models excluding either of these two factors. The deviance statistics show that the omission of co-occurrence frequency as a predictor does not give a significantly worse fit ($X^2 = 3.609$, $df = 2$, $p = 0.165$), but the omission of the relatedness predictor does ($X^2 = 8.148$, $df = 2$, $p = 0.017$).

We then continued fitting four simpler models, each without the co-occurrence predictor and without any of the four remaining predictors. The results show that lag size can be omitted from the model without leading to a significantly worse fit ($X^2 = 0.005$, $df = 1$, $p = 0.943$), but none of the other predictors (target frequency: $X^2 = 3.988$, $df = 1$, $p = 0.046$; target length: $X^2 = 14.007$, $df = 1$, $p = 0.000$; trial: $X^2 = 5.980$, $df = 1$, $p = 0.014$) can be omitted. The coefficients of this final model with the four predictors are displayed in Table 6.

There is a negative relationship between trial and reaction times, indicating that the participants tend to become slower towards the end of the experiment. Similarly, there is a negative relationship between reaction times and the frequency of the target, indicating that the participants respond to high-frequency targets faster than to low-frequency targets. The positive coefficients for relatedness and target length indicate that responses to antonyms are faster than those to unrelated adjectives, and that responses to longer targets are slower than responses to shorter targets.

Table 6. Final model

	coefficient estimate	standard error	*t*
trial	−0.0002	0.0001	−2.52*
relatedness	0.0270	0.0127	2.13*
target frequency (log)	−0.0134	0.0068	−1.98*
target length	0.0124	0.0033	3.80*

6. Discussion

As stated in the introduction, previous research using experimental methods as well as corpus techniques show that the relation of antonyms has special status in terms of semantic priming effects and in terms of co-occurrence in text and discourse and that some very frequent and frequently co-occurring

adjectives stand out as excellent members of the category. Antonyms prime each other more often than other related words and they co-occur in sentences significantly more often than other words. It has also been shown that the most strongly co-occurring antonyms in text are also the ones that people judge to be excellent antonyms in judgement experiments. In elicitation experiments these adjectives elicit only one or a couple of antonyms, which we interpret as an indication of 'goodness of antonymy' both in terms of semantic relatedness and lexical association. On the basis of their priming experiments, Spence and Owens point out that frequency of co-occurrence is a function of association strength, and unlike the data used in Charles and Miller (1989), Justeson and Katz (1991, 1992), Paradis et al. (2009) and Willners and Paradis (2010), their co-occurrence data extend beyond the borders of the sentence. They report that priming effects do not seem to decline until some 200 words beyond the stimulus word. However, there are also studies with results that point in a different direction. Estes and Jones (2009) investigated whether frequency of co-occurrence in large samples of written language provides better estimates of prime-target familiarity. They, too, used both adjacent words and global co-occurrence in text, using Latent Semantic Analysis (LSA) for the latter type of co-occurrence. They found no crucial priming effects related to co-occurrence frequency, neither locally nor globally.

In spite of the difference of semantic relatedness across the above studies and the different types of experiments using co-occurrence frequency, the contradictory results of Estes and Jones (2009) led us to look more closely into whether goodness of antonymy is a matter of strength of relatedness or lexical association and to what extent frequency of co-occurrence plays a role for the status of strongly canonical antonyms. In contrast to our own, more semantically oriented previous experiments, we designed this experiment as a lexical recognition task in order to put the spotlight on the lexical side of the matter. We argued that if it is the case that antonyms which co-occur often in text and discourse are considered better antonyms than antonyms that do not co-occur often, then we would expect such form-meaning pairings to be more strongly entrenched in memory than less frequently co-occurring pairs of form-meaning pairings. As a result, the presentation of one member of a frequently co-occurring pair will facilitate the recognition of the other member. This will not be the case if the members of a pair do not co-occur often.

In order to investigate this state of things, we selected pairs of adjectives with varying degrees of co-occurrence frequencies for the BNC. Half of these pairs were antonyms, and the other half were not related in meaning.

We carried out a visual decision experiment to see whether we would find evidence of priming from one member of a pair to the other. The results showed that the recognition of antonymic targets was indeed facilitated by their primes, which was not the case for the unrelated pairings. We also found a facilitatory effect of target word frequency on reaction times: frequent targets were recognized faster than less frequent targets. Crucially, however, we did not find that frequently co-occurring antonyms, such as *horizontal* and *vertical*, primed each other more than less frequently co-occurring antonyms, such as *distant* and *near*. This means that when participants see *horizontal*, they lower the threshold for *vertical*. This effect is not due to the fact that *horizontal* and *vertical* co-occur frequently but because they are semantically related. The facilitation of the target word also happens in the case of less frequently co-occurring antonyms such as *distant* and *near*. Facilitation is thus not likely to be a consequence of lexical association, and it cannot be attributed to familiarity either, but has a semantic basis. Similarly, in the case of the unrelated test items, we found no priming effect and no facilitating effect due to frequency of co-occurrence either.

In our experiment we manipulated lag size between the prime and the target, which is a different procedure from Estes and Jones' (2009) experiment on co-occurrence and compound nouns. They used Stimulus Onset Asynchronies (SOA) to investigate different types of priming and found that both associative priming and semantic priming are observed at short (i.e. <300 ms) and intermediate SOAs (approximately 300 ms–800 ms). Associative priming continues to increase in magnitude across longer SOAs (i.e. $\geq 1{,}000$ ms), whereas semantic priming tends to dissipate at those later SOAs (e.g. Perea and Rosa 2002b). Instead of using SOAs, we manipulated the number of words in between stimuli. The stimuli were presented with either zero, one or two nonsense words between the prime and the target. The reason why we varied the distance was mainly to avoid testing fatigue and monotony and to prevent the participants from developing experiment strategies.

The analysis started with an initial model containing all predictors, item frequency, item length, lag size (0, 1 or 2 words), and the joint effects (i.e., the interaction plus the main effects) for relatedness and co-occurrence frequency. The outcome of the analysis suggests significant effects of trial, target length, and target frequency, but not of relatedness, lag size, co-occurrence frequency nor of their interaction. It would be interesting to repeat our experiment using varying SOAs instead of intervening items.

Other potentially confounding variables may be participants' level of word knowledge. Yap et al. (2009) isolate various variables in order to be able to show a more nuanced picture and in order to better explain the relation between priming effects and word frequency. Through a multi-stage model of lexical processing consisting of a lexical (perceptual, orthographic in this case) module and a semantic (cognitive) module, they measure processing differences across participants on the basis of their level of vocabulary knowledge. They show that the joint effects of semantic priming and word frequency are critically dependent on the participants' level of vocabulary knowledge. Yap et al. also show that semantic priming and word frequency do not always interact. Participants with less vocabulary knowledge show larger priming effects than participants with higher vocabulary knowledge. The priming effects among the former group of participants are particularly pronounced for low-frequency targets. They argue that the result is consistent with the idea of a flexible lexical processing system in which participants' performance is optimized by relevant contextual information. In contrast, the lexical processing system of participants with higher vocabulary knowledge appears to be more modular in nature, i.e. the prime provides a head-start that is independent of how difficult the target is. We do not think that participants' vocabulary knowledge plays a role in our data, because the participants are all well-educated speakers with university education and all the test items are common words in English, occurring at a rate of more than 1,000 times in the corpus.

Priming is typically attributed to strength of lexical association and semantic relatedness. Frequency effects are obtained: frequent words are recognized faster than rare words. Similarly, if words preceded by semantically related primes are recognized faster than unrelated primes, we have a semantic effect. This distinguishes between association priming, which is caused by lexical associative strength, and semantic priming, which is due to semantic similarity. There are a number of theoretical models available for the explanation of lexical priming in the literature: association models such as spreading activation (e.g. Anderson 1983), and expectancy models (e.g. Becker 1980). There are also more semantically oriented models such as the distributed representation model (e.g. Plaut and Booth 2000), which assumes priming to occur when words have overlapping patterns of activation of semantic features represented in different parts of the brain. Moreover, there is the compound cue model (Ratcliff and McKoon 1988), which posits that a prime forms a cue which is matched against long-term memory, in which case the prime–target relation is a result of the extent

to which they are associated in memory, or in Ratcliff and McKoon's terminology, the extent of the familiarity of the pairing. In their model, co-occurrence in text has the status of being the best predictor of strength of familiarity (McKoon and Ratcliff 1992).

This view has been challenged by Estes and Jones (2009) who showed that lexical priming also occurs among unassociated, dissimilar and unfamiliar concepts such as *horse* and *doctor*. Such priming is said to occur when a prime word can be easily integrated with the meaning of a target word to create a unitary representation. It was also shown that integrative priming was different from associative and semantic priming but comparable to them in terms of prevalence across the participants as well as magnitude within participants. Estes and Jones (2009) argue that this finding constitutes a challenge to models such as spreading activation, distributed representation, expectancy, episodic retrieval and compound cue models, and they suggest that it can be explained by a role activation model of relational integration. In spite of the fact that the test items in Estes and Jones (2009) were compound concepts, unlike our semantically related antonyms, the result of that experiment is similar to ours in that they did not obtain any co-occurrence frequency effects.

7. Conclusion

This study has two important results. It confirms the hypothesis and previous findings that antonymic targets are facilitated by their primes, but it does not confirm the hypothesis that frequency of co-occurrence facilitates word recognition, either for antonyms or for unrelated adjectives. This means that there is a relatedness effect but no co-occurrence frequency effect, which in turn means that priming cannot be attributed to lexical association. The prime-target effect we obtain is a semantic effect, indicating that conceptual opposition is the *cause* of lexical relation rather than the other way round the *effect* of the lexical relation. This piece of evidence lends support to a conceptual rather than a lexical approach to antonymy.

Appendix A: Experimental items with co-occurrence frequencies

Antonyms		*Unrelated*	
large-small	7.81	little-nice	6.30
high-low	7.62	foreign-prime	6.25
bad-good	7.39	british-nuclear	6.03
female-male	7.33	other-sufficient	5.70
old-young	7.15	medical-national	5.20
long-short	7.12	great-private	5.05
negative-positive	7.09	appropriate-relevant	5.04
right-wrong	6.60	effective-simple	5.02
poor-rich	6.59	big-real	4.97
different-similar	6.51	early-popular	4.93
cold-hot	6.48	current-total	4.89
false-true	6.12	legal-personal	4.88
horizontal-vertical	5.97	financial-serious	4.88
strong-weak	5.92	general-late	4.86
dark-light	5.86	major-rural	4.84
domestic-international	5.83	necessary-proper	4.83
hard-soft	5.72	immediate-political	4.81
difficult-easy	5.72	natural-public	4.80
spoken-written	5.70	broad-economic	4.79
permanent-temporary	5.68	annual-royal	4.65
dry-wet	5.60	heavy-main	4.26
closed-open	5.57	democratic-leading	4.23
active-passive	5.56	deaf-special	4.22
practical-theoretical	5.54	fresh-whole	4.20
alive-dead	5.53	likely-powerful	4.17
new-used	5.45	available-safe	4.16
front-rear	5.35	social-vast	4.14
empty-full	4.92	black-flat	4.14
acute-chronic	4.88	formal-particular	4.13
absolute-relative	4.88	familiar-important	4.11
multiple-single	4.88	complete-human	4.09
narrow-wide	4.87	possible-straight	4.01
cheap-expensive	4.86	common-useful	3.97
fast-slow	4.85	deep-pink	3.95
global-local	4.80	sharp-white	3.95
classical-modern	4.78	criminal-reasonable	3.47
thick-thin	4.72	fine-usual	3.09
absent-present	4.55	responsible-standard	3.04
liquid-solid	4.44	interesting-urban	2.83
clean-dirty	4.43	blue-loose	2.71
final-initial	4.42	fair-fat	2.30

happy-sad	4.36	previous-terrible	2.08
mild-severe	4.25	raw-sweet	2.08
basic-complex	4.23	honest-pleasant	1.95
abstract-concrete	4.22	recent-thorough	1.95
falling-rising	4.16	quick-tight	1.79
cool-warm	4.16	decent-essential	1.79
beautiful-ugly	3.97	bitter-radical	1.61
compulsory-voluntary	3.93	brave-pretty	1.39
rough-smooth	3.53	calm-correct	1.10
normal-strange	3.14	crude-free	1.10
gradual-sudden	3.14	careful-tiny	1.10
clever-stupid	3.09	famous-wise	1.10
clear-vague	3.09	dear-mad	0.69
boring-exciting	2.83	angry-central	0.69
healthy-ill	2.77	subtle-violent	0.69
friendly-hostile	2.71	busy-plain	0.69
cruel-gentle	1.95	guilty-proud	0.69
odd-regular	1.79	causal-genuine	0.69
funny-tragic	1.61	blind-silly	0.69

References

Anderson, John R.
1983 A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior* 22: 261–295.

Baayen, R. Harald
2008 *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press.

Becker, Curtis. A.
1979 Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 5: 252–259.

Becker, Curtis. A.
1980 Semantic context effects in visual word recognition: an analysis of semantic strategies. *Memory and Cognition* 8: 493–512.

Charles, Walter G. and George A. Miller
1989 Contexts of antonymous adjectives. *Applied Psycholinguistics* 10(3): 357–375.

Croft, William and Alan Cruse
2004 *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Estes, Zachary and Lara Jones
2009 Integrative priming occurs rapidly and uncontrollably during lexical processing. *Journal of Experimental Psychology* 138(1): 112–130.

Fellbaum, Christiane
1995 Co-occurrence and antonymy. *International Journal of Lexicography* 8(4): 281–303.
Gainotti, Guido, Gabriele Miceli, M. Caterina Silveri and Giampipero Villa
1983 The production of morphological and lexical opposites in aphasia. *Neuropsychologia* 21(6): 693–697.
Herrmann, Douglas J., Roger Chaffin, Margaret P. Daniel and R. S. Wool
1986 The role of elements of relation definition in antonym and synonym comprehension. *Zeitschrift für Psychologie* 194: 133–153.
Jones, Steven
2002 *Antonymy: a corpus-based perspective*. London: Routledge.
Jones, Steven
2007 'Opposites' in discourse: a comparison of antonym use across four domains. *Journal of Pragmatics* 39(6): 1105–1119.
Jones, Steven and M. Lynne Murphy
2005 Using corpora to investigate antonym acquisition. *International Journal of Corpus Linguistics* 10(3): 401–422.
Jones, Steven, Carita Paradis, M. Lynne Murphy and Caroline Willners
2007 Googling for 'opposites': a web-based study of antonym canonicity. *Corpora* 2(2): 129–154.
Justeson, John S. and Slava M. Katz
1991 Co-occurrences of antonymous adjectives and their contexts. *Computational Linguistics* 17: 1–19.
Justeson, John S. and Slava M. Katz
1992 Redefining antonymy: the textual structure of a semantic relation. *Literary and Linguistic Computing* 7: 176–184.
McKoon, Gail and Roger Ratcliff
1992 Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18: 1155–1172.
Miller, George A. and Christiane Fellbaum
1991 Semantic networks of English. *Cognition* 41: 197–229.
Muehleisen, Victoria and Maho Isono
2009 Antonymous adjectives in Japanese discourse. *Journal of Pragmatics* 41(11): 2185–2203.
Murphy, Gregory L.
2002 *The big book of concepts*. The MIT Press: Cambridge (MA).
Murphy, Gregory L. and Jane M. Andrew
1993 The conceptual basis of antonymy and synonymy in adjectives. *Journal of Memory and Language* 32: 301–319.
Murphy, M. Lynne
2003 *Semantic relations and the lexicon*. Cambridge: Cambridge University Press.
Murphy, M. Lynne and Steven Jones

2008 Antonymy in children's and child directed speech. *First Language*, 28(4): 403–30.

Murphy, M. Lynne, Carita Paradis, Caroline Willners and Steven Jones
2009 Discourse functions of antonymy: a cross-linguistic investigation. *Journal of Pragmatics* 41(11): 2159–2184.

New, Boris, Ludovic Ferrand, Christophe Pallier and Marc Brysbaert
2005 Re-examining the word length effect in visual word recognition: new evidence from the English lexicon project. *Psychonomic Bulletin and Review* 13: 45–52.

Paradis, Carita and Caroline Willners
2007 Antonyms in dictionary entries: methodological aspects. *Studia Linguistica* 61(3): 261–277.

Paradis, Carita and Caroline Willners
2011 Antonymy: from convention to meaning-making. *Review of Cognitive Linguistics* 9(2): 367–391.

Paradis, Carita, Caroline Willners and Steven Jones
2009 Good and bad opposites: using textual and experimental techniques to measure antonym canonicity. *The Mental Lexicon* 4(3): 380–429.

Paradis, Carita
2010 Good, better and superb antonyms: a conceptual construal approach. *The Annual of Texts by Foreign Guest Professors*. Charles University in Prague, Faculty of Philosophy and Arts.

Plaut, David C. and James R. Booth
2000 Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review* 107: 786–823.

Perea, Manuel and Eva Rosa
2002a The effect of associative and semantic priming in the lexical decision task. *Psychological Research* 66: 180–194.

Perea, Manuel and Eva Rosa
2002b Does the proportion of associatively related pairs modulate the associative priming effect at very brief stimulus-onset asynchronies? *Acta Psychologica* 110: 103–124.

Ratcliff, Roger and Gail McKoon
1988 A retrieval theory of priming in memory. *Psychological Review* 95: 385–408.

Samson, Dana, Catherine Connolly and Glyn W. Humphreys
2007 When "happy" means "sad": neuropsychological evidence for the right prefrontal cortex contribution to semantic processing. *Neuropsychologica* 45(5): 896–904.

Schneider, Walter, Amy Eschman and Anthony Zuccolotto
2001 *E-prime user's guide*. Pittsburgh: Psychology Software Tools, Inc.

Spence, Donald P. and Kimberly C. Owens
1990 Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research* 19: 317–330.

Singer, Judith and John Willett
2003 Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press.

Söderpalm, Ewa
1979 Speech errors in normal and pathological speech. *Travaux de l'Institut de Linguistique de Lund* 14.

Taylor, John R.
2003 *Linguistic Categorization*. Oxford: Oxford University Press.

Tribushinina, Elena
(in press) Spatial adjectives in Dutch child language: towards a usage-based model of adjective acquisition. In Carita Paradis, Jean Hudson, and Ulf Magnusson (eds.), *The Construal of Spatial Meaning: Windows into Conceptual Space*. Oxford: Oxford University Press.

Varley, Rosemary
1991 Reference, sense and antonymy in the assessment of lexical semantic abilities in aphasia. *Aphasiology* 5(2): 149–170.

Willners, Caroline
2001 Antonyms in Context. A Corpus-based Semantic Analysis of Swedish Descriptive Adjectives. *Travaux de l'Institut de Linguistique de Lund* 40.

Willners, Caroline and Carita Paradis
2010 Swedish opposites: a multi-method approach to goodness of antonymy. In Petra Storjohann (ed.), *Semantic relations: Theoretical and practical perspectives*, 15–48. Amsterdam: John Benjamins.

Yap, Melvin, Chi-Shing Tse and David Balota
2009 Individual differences in the joint effects of semantic priming and word frequency revealed by RT distributional analyses: The role of lexical integrity. *Journal of Memory and Language* 61: 303–325.

Subject index