Article:

Divjak, D.S. (2017) The role of lexical frequency in the acceptability of syntactic variants: evidence from that-clauses in Polish. Cognitive Science, 41 (2). pp. 354-382. ISSN 1551-6709

https://doi.org/10.1111/cogs.12335

The role of lexical frequency in the acceptability of syntactic variants: evidence from *that*-clauses in Polish

(Author list and contact information removed for the purpose of preserving anonymity)

Abstract

A number of studies in both the generative and usage-based traditions report that frequency is a poor predictor of acceptability in morphology and syntax, in particular at the lower end of the frequency spectrum. Because acceptability judgments provide a substantial part of the empirical foundation of dominant linguistic traditions, understanding how acceptability relates to frequency, one of the most robust predictors of human performance, is crucial.

The relation between low frequency and acceptability is investigated using data on the distribution of infinitival and finite *that*-complements in Polish. Polish verbs exhibit substantial subordination variation and for the majority of verbs taking an infinitival complement, the *that*-complement occurs with low frequency (< 0.66 ipm) in a 1.5 billion word corpus. This does, however, not affect its acceptability.

A mixed effects ordinal regression model shows that the verb's morphological transparency contributes significantly to its acceptability in a syntactic alternative: the more recognizably the verb is related to other words, the higher the acceptability rating. This is true in particular for verbs that are in the bottom quartile (0.4 to 6 ipm) for unigram frequency. The probability of encountering a verb in a syntactic alternative conditioned on the verb explains a further significant share of the variance: an increase in the degree to which a verb relies on the *that*-construction correlates positively with an increase in the acceptability of the verb in the construction. This effect is most strongly observed for verbs with unigram frequencies in the middle quartiles (10 to 140 ipm).

Acceptability judgments are thus not based on n-gram frequency but on configurations of internally structured exemplars, i.e. on higher-order knowledge known as schemata or rules. The findings also show that speakers cannot reliably formulate direct hypotheses about the acceptability of such configurations until sufficient evidence about the core component has accumulated.

Keywords

acceptability ratings, association measures, comprehension, conditional probability, frequency effects, prediction, probabilistic grammar, Reliance, Surprisal, usage-based theories of language

1. Introduction[1]

Frequency is among the most robust predictors of human performance (Hasher and Zacks 1984). One of the first studies to address the influence of word frequency on word recognition was Cattell (1886) who showed that higher frequency words are recognized more quickly than lower frequency words. Since then, a large number of studies have investigated the extent to which different forms of linguistic behaviour would be experience-driven, and evidence has been found for a range of phenomena, from processing single words to acquiring knowledge of the sets of verbs that are used in complex argument structure constructions (for overviews see Sedlmeier and Betsch 2002; Ellis 2002; Diessel 2007; Gries and Divjak 2012; Divjak and Gries 2012).

The fact that speakers' language systems are sensitive to frequencies of occurrence in language use has been accommodated in a different way by generative and usage-based linguists. While generative linguists consider frequency to be a performance factor which influences linguistic processing rather than the inventory of stored entities (Chomsky 1957; Newmeyer 2003) or relegate these effects to the mental lexicon (Ullman and Walenski 2005), usage-based linguists claim that frequency, as proxy of experience, plays a central role in the emergence and entrenchment of linguistic units: surface distributions contain the necessary information to build up adequate mental linguistic representations. Surface token frequency motivates learning through repetition: the token is the instance that is repeated and subsequently learned. The more often a pattern is experienced, the easier it becomes to access and use (see articles in Bybee and Hopper 2001). Given the crucial role of repetition from which frequency arises, the process of language acquisition is claimed to be probabilistic in nature, with earlier and higher proficiency expected for the more frequent items. These highly-frequent formulas form templates that

[1] I would like to thank Antti Arppe, Neil Bermel, Catherine Caldwell-Harris, Jane Klavan, Natalia Levshina and the reviewers for Cognitive Science (Harald Baayen, Stefan Frank and Amir Zeldes) for detailed comments on a previous version of this paper.

gradually develop into distinct (potentially only low-level) schemas or types through the categorization of further exemplars, thus enabling the emergence of a grammar from the ground up (Bybee 2006 and elsewhere). The resulting mental representation or grammar would thus be probabilistic in nature (Bod et al. 2003), as a wide range of other cognitive processes appear to be (Rao et al. 2002). Although the presence of frequency effects is not in itself sufficient to warrant adopting a probabilistic view on language, it does signal that the basic building blocks of probability theory, occurrence frequencies, may be exploited (Bod et al. 2003: 3).

Within the Generative School, the type is not arrived at by token-repetition; instead, the type exists in Universal Grammar, and manifests itself as tokens in language use. The generative understanding of type:token is thus driven top-down, and token frequency is not permitted to alter the grammar that is governed by Universal Grammar. Token frequency has therefore long remained largely irrelevant to Generative models of language, except in parameter setting (Yang 2004). Yet, as Tomasello (2007:282) puts it: “[t]oday, very few linguists would seriously deny the existence of frequency effects in language. The real argument within linguistics is how far these effects go”.

Usage-based linguists who propose single-system models predict frequency effects for all linguistic units: simple and complex, lexical and grammatical. Frequency effects have typically been observed using longitudinal corpus data or in the relation between corpus counts and behavioural data. One area where frequency seems to have run into problems is that of acceptability or grammaticality judgments. In accord with usage-based theory one would expect that “grammaticality or acceptability judgments are heavily based on familiarity, that is, the speaker’s experience with language in use. Sequences of linguistic units that are of high frequency or resemble sequences of high frequency will be judged more acceptable than those that are of low frequency or do not resemble frequently used structures” (Bybee and Eddington 2006: 349). A number of studies in both the generative and usage-based traditions have, however, confirmed the existence of a grammaticality-frequency discrepancy, if not a gap (Kempen and Harbusch 2005; 2008), for acceptability ratings: corpus frequencies are poor predictors for off-line acceptability ratings, in particular at the lower end of the frequency spectrum, in morphology and syntax (Keller 2003; Kempen and Harbusch 2005/2008; Arppe and Järvikivi 2007; Divjak 2008; Bader and Häussler 2009; Bermel and Knittl 2012a/b; but see the opposite tendency in the results of Lapata et al. 1999 for adjective-noun combinations). This has strengthened generativists in their belief that "simple frequency data" could and should be ignored in theoretical linguistic analyses.

Yet, that low frequencies would cause problems when used to predict acceptability is expected on a probabilistic approach: prediction mechanisms cannot make reliable inferences without sufficient data (Evert 2005: 133, 166). Curiously, a large proportion of elements that are nevertheless successfully acquired have a very low frequency of occurrence (cf. Zipf 1949; Mandelbrot 1965; Baayen 2001; Evert 2005). To become a fully competent speaker of a language, the learner must thus be able to overcome the challenges posed by the Zipfian distribution. In Generative circles, this is achieved by grammars that project beyond input data (Chomsky 1975; Legate and Yang 2002 for a quantitative treatment within a generative framework; Ramscar and Yarlett 2007 for an account in terms of discrimination learning); in usage-based frameworks, learners need to collect the information they need to deal efficiently with linguistic expressions of the entire frequency range.

The current study therefore investigates the extent to which usage, and a speaker's experience of it, contributes to the acceptability of complex lexico-syntactic structures. This is achieved by relating acceptability ratings for verbs that occur with low frequency in *that*-constructions with a range of variables capturing information relating to the (co-)occurrence, morphology, and semantics of the verbs and the *that*-construction. The results help determine, on the one hand, how far usage frequencies take the learner or speaker in knowing whether or not to accept a potential syntactic alternant and thus to avoid (overgeneralization) errors. On the other hand, the findings reveal what type of frequency-based measures, as proxies for different summaries of experience with language, yield the most accurate predictions about the availability and acceptability of an alternative syntactic phrasing. More generally, the findings refine the understanding of experience in the current usage-based model by considering the effects that different types of knowledge, distilled from usage frequencies, have on the mental representation of larger syntactic constructions. Furthermore, given that more than two thirds of the lexico-syntactic combinations included in this study occur with very low frequency, i.e. less than 0.66 times per million words, the findings shed light on the question of whether memory traces are available for exemplars at the lower end of the frequency range and whether there is a threshold below which frequency would not be operational.

2. Methods[2]

The relation between frequency and acceptability is investigated on the basis of the distribution of infinitival and finite *that*-complements in Polish. This phenomenon has received ample attention within Generative frameworks under the headers of (Subject/Object) Control and (Subject) Obviation for a range of languages, including Polish (Bondaruk 2004; Dziwirek 1998; Dziwirek 2000; Przepiórkowski and Rosen 2005, among others). Control verbs in Polish differ with respect to whether they allow, require or resist the presence of the complementizer *żeby* and the meaning of the verb does not affect this (Bondaruk 2004: 208). This begs the question of how learners can acquire a system without a semantic or functional motivation. In order to shed light on this issue, data from all verbs that are known to occur with an infinitive complement will be used: testing a natural category in its entirety facilitates exploring the frequency hypothesis in a naturalistic setting, typically encountered by learners and speakers of a language.

An example illustrates that it is possible to use both infinitival and finite complements in co-referential sentences with *decide* as the main verb, such as (1), but not in sentences such as (2) with *want* as the matrix verb. The question is: how do speakers know when a *that*-alternative is and is not available?

(1)	Zdecydował	wyjechać.
	Decided PF.IND.PAST.MASC.3SG	leave PF.INF
	He decided to leave.	

[2] I would like to thank Dr Agnieszka Będkowska-Kopczyk, Dr Władysław Chłopicki, Łukasz Degórski, Adrianna Jakóbczyk, Dr Agnieszka Mikołajczuk, Monika Prokopczuk, Dr Anna Słoń, Prof Elżbieta Tabakowska, em. Prof. Andrzej Uggla†, Torkel Uggla, Dr Kris Van Heuckelom and Dr Chris Wiesen for their assistance with different parts of the data collection. Amy Baddeley, MSc in Statistics, and Jean Russell, Chartered Statistician, provided statistical consultancy. I gratefully acknowledge the financial support of the Belgian American Educational Foundation (B.A.E.F.), the Erasmus program, the Prokhorov Foundation and the University of Sheffield. The research received ethical approval from the UNC-CH (USA).

Zdecydował, że wyjedzie.

Decide PF.PAST.MASC.3SG that leave PF.IND.NON_PAST.3SG

He decided that he would leave.

(Grzegorczykowa 2006: 83)

(2) Chciał wyjechać.

Want IMPF.IND.PAST.MASC.3SG leave PF.INF

He wanted to leave.

*Chciał, że wyjedzie.

Want IMPF.PAST.MASC.3SG that leave PF.IND.NON_PAST.3SG

He wanted that he would leave.

The *that*-clause alternative is likewise available for some non-coreferential infinitival clauses; compare here (3) with (4). Note that in this type of sentence the main clause contains a dative objective rather than an accusative object (Dziwirek 1998) and that the preferred complemetizer is *żeby*.[3]

(3) Profesor kazał mi powtórzyć kurs.

Professor order IMPF.IND.PAST.MASC.3SG me DAT.1SG repeatPF.INF module.

The professor instructed me to repeat the module.

Profesor kazał mi żeby-m powtórzył kurs

Professor ordered me that-1SG repeated PF.IND.PAST.MASC.SG module

The professor told me that I should repeat the module

[3] The Polish *that*-complement construction comes in three main forms, i.e. *że* + finite verb, *żeby* + infinitive or *żeby* + past tense verb (or strictly speaking a conditional, since the *–by* attached to *że* is a conditional particle). By and large, the choice of conjunction is determined by the meaning of the main verb while depending to some extent on the reality and controllability of the subordinate clause: if the state of affairs is a fact that is (presented as) certain to happen, *że* is an adequate conjunction. The precise form of the verb in the *żeby*-clause mainly depends on co-referentiality between the subjects of the main clause and the subordinate clause: if there is co-referentiality, *żeby* + infinitive may be used. Note that, in some cases, a form of the demonstrative pronoun *to* precedes the clause introduced by *żeby*. Apart from the stylistically neutral conjunctions *że* and *żeby*, bookish *iż* can replace *że* while neutral *by*, bookish *ażeby*, colloquial *coby* and obsolete *iżby* can replace *żeby*.

(Swan 2002: 260)

(4) Marek dał mi poprowadzić swój samochód.

Marek let IMPF.IND.PAST.MASC.3SG me DAT.1SG drive PF.INF his car

Marek let me drive his car.

*Marek dał mi żebym poprowadził swój samochód.

*Marek dał mi żeby poprowadzić swój samochód.

(Bondaruk 2004:207)

In Polish, 122 verbs are tagged in dictionaries (Bańko 2000; Polański 1988-1992) as taking an infinitive complement construction; 95 of these are morphologically unrelated to each other and are still productively used with an infinitive complement (for a complete list with English translations see Appendix 1). What holds this category together is not immediately apparent: there is no clear shared semantic meaning that would unite the category, and neither do any substantial semantic sub-fields emerge (Bondaruk 2004: 208). Dictionaries list 41 of these verbs as taking an infinitive as well as a *that*-clause (but see Appendix 1 and Section 2.3 for actual usage data); for this subgroup no shared semantic core has been identified either.

2.1. Experimental set-up

The data were collected using an off-line acceptability rating paradigm (Sprouse 2013). Although reservations towards the use of acceptability or grammaticality judgments have been expressed , these judgments provide a substantial part of the empirical foundation of nearly every area of linguistics, and nearly every type of linguistic theory. They have been used particularly intensively in contemporary generative syntax (Schütze 1996; Cowart 1997; Sprouse and Almeida 2013) and in usage-based language acquisition research (Ambridge 2011). It is therefore crucial to understand how these judgments relate to one of the core concepts in cognitive science and contemporary linguistic theory: frequency.

Acceptability ratings are considered off-line measurements, i.e. they provide information about language knowledge that is available after the initial stages of processing, which are typically operationalized as the first 300 ms after onset. Off-line tasks reduce the need to rely on routinized structures, thus allowing speakers to consult “the system”, i.e. the abstraction arising from exposure to language on the usage-based approach (be it stored or created on the fly). In other words, off-line tasks provide information about the final outcome of language processing, about what is permanent rather than ephemeral (Kaiser 2013: 137). Yet language processing constraints do enter the picture (Sprouse 2008; Staum et al 2010; Hofmeister et al. 2012) because judging an utterance involves trying to comprehend it and could also involve the attempt to generate the utterance in question with an utterance being judged grammatical if it can be generated, and ungrammatical if not. Schütze (1996) makes the assumption that the judgment process is additionally influenced by a number of subject-related and task-related components that are not necessarily involved in language processing.

Participants

285 undergraduate students of English/German in Poland, all native speakers of Polish, participated in the study In addition to acceptability ratings, they provided basic "demographic" information, such as year of birth, native language, parental education, major subject and handedness, which is known to influence aspects of linguistic behavior, in particular sentence processing (Townsend et al. 2001; Cowart 1989).

Materials

The 285 experimental sentences were (shortened) authentic *that*-sentences extracted from the newspaper section within the PELCRA reference corpus of Polish (http://korpus.ia.uni.lodz.pl/); in case no *that*-sentences were attested, some were created from infinitive sentences found in the same sub-corpus using the most likely form of the *that*-clause, as judged by 5 native speakers of Polish. In order to neutralize lexical effects of any items other than the verb (Schütze and Sprouse 2013: 39), three different lexicalizations were provided for each of the 95 verb**that*-construction combinations. 25 filler and 10 benchmark sentences were adapted from authentic sentences extracted from newspapers to be comparable to the experimental sentences in plausibility, complexity and length and to instantiate grammaticality levels ranging from -2 to +2. Overall, the ratio between experimental sentences and fillers was 1:9 in the survey, and within each block of 8 sentences, only 1 was an experimental sentence.

In each questionnaire, 5 of the experimental sentences (each with a different verb) and 25 fillers were randomly assigned to 5 blocks and then shuffled within blocks. The first block was preceded by a block of 5 benchmark sentences; the last block was followed by 5 benchmark sentences. Participants were asked to indicate “how Polish this sentence sounds” on a 5-point Likert scale with the following instructions:

Very strange, unnatural Polish		**OK Polish, could be heard**		**Natural Polish**
-2		**0**		**2**

Participants were ensured there were no right or wrong answers and were asked not to revisit previous answers. For the filler scores, the mean was taken across all 25 sentences for every rater.

Procedure

The participants filled out the survey in a classroom setting. Instructions and examples were provided in English on the questionnaires, and were translated into Polish by the experimenters. Participation was voluntary: no course credit or financial compensation was provided for participating in the study, and subjects were told they could quit at any time. No time limit for completion was set, but all participants returned the questionnaire after 10 to 15 minutes.

2.2. Explanatory variables

The extent to which experience contributes to the acceptability of complex lexico-syntactic structures is assessed by relating acceptability ratings for *that*-constructions with a range of variables that capture probabilistic, morphological, and semantic information.

2.2.1 Frequency information

Previous studies set up to show that surface distributions contain the necessary information to build up adequate mental linguistic representations have settled for raw frequencies, i.e. simple corpus counts of verbs or constructions as predictors of behaviour (but see Theakston 2004:29 for a remark that the frequency of verbs in constructions may in fact be more revealing). Yet human beings are capable of much more intricate forms of statistical learning (Saffran 2003) and what is learned or acquired by probabilistic means is not strictly proportional to the stimulus: 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). Capturing this ability more adequately might push the lower bound for the usefulness of corpus frequencies further down.

Frequency data was obtained from the 1.5 billion word version of the NKJP [nkjp.pl], the Polish National Corpus. All texts, with the exception of older prose and a small number of contemporary prose texts, were created from the 1990s onwards. The corpus is fully parsed with an overall tagging accuracy of up to 98% (Adam Przepiórkowski 2007, p.c.). Data on the *that*-construction for the 95 verbs studied were extracted by means of regular expressions written for the stand-alone version of Poliqarp (Przepiórkowski 2004; Janus and Przepiórkowski 2006), and the samples were manually cleaned (see section 2.2.1.2).

For all frequency-based variables, the natural logarithm was taken since the log(token) is known to correlate with psycholinguistic processes, notably with word recognition times (see Howes and Solomon 1951 for a first report).[4]

2.2.1.1 The frequency of the word in the corpus

First, the unigram frequency of the verb in the corpus was recorded; this measure of overall frequency was included to give an idea of how likely raters would be to know the verb in question.[5] Within the complete sample, occurrences ranged from 0.196 to 6220 ipm. Within the subsample of verbs attested in the *that*-construction in the corpus, occurrences ranged from 0.393 to 1516.32 ipm.

Raw frequencies of the rate with which both the infinitive and the *that*-complement construction were encountered were also collected. The infinitival and *that*-constructions differ in terms of token and type frequency distribution. The infinitival complement construction (defined as finite verb followed by infinitive with up to 3 words intervening within clause boundary) occurs 15,814,680 times while the *that*-complement construction (defined as finite verb followed by *that*

4 Some of the verbs have a value of 0 for occurrence in the *that*-construction, and a logarithm cannot be calculated. Since measures for dealing with zero frequencies (Brysbaert & Diependaele 2013) were found to distort the data, these observations were excluded from the analysis. Using the binomial distribution, the corpus size needed to encounter these verbs at least once in a *that*-clause is estimated at 15 billion words, that is 10 times larger than the largest corpus currently available.

5 Dispersion was not calculated because the corpus is not available in a format that would allow computing measures as those outlined in Gries (2008, 2010, with corrections in Lijffijt & Gries 2012). It would be interesting to see whether dispersion improves the correlation between frequencies and acceptability ratings the way it improves response times in word recognition studies. Yet given that the vast majority of the texts included in the corpus are rather short, that the construction of interest is low frequent (with fewer than 0.66 ipm for more than two thirds of the data) and that dispersion is highly correlated with word frequency (McDonald and Shillcock 2001), little variation in dispersion is expected.

no more than 3 words apart within a sentence) is used 4,137,026 times. The nearly 4 times more frequent infinitival construction is attested with 122 verbs, among which there are some highly frequent verbs such as the auxiliary verb *być* ('be') in the future tense, and modal verbs *móc* ('can, may'), *mieć* ('have') and *musieć* ('have to'). This gives the infinitival construction a lower type but higher token frequencies, which promotes verb-specific construction learning (Goldberg et al. 2004). The less frequent *that*-construction is considerably more flexible and follows a wide range of verbs and even nouns, thus exhibiting a higher type but lower token frequency. This makes it more likely that a general category is formed for the *that*-construction that is readily available to extend to new items.

2.2.1.2 Contextual frequencies: the frequency of the verb in the construction

These individual verb frequencies can be contextualized.[6] First, the contextual or joint frequency of the word in the infinitive complement construction was determined. This measure was included as a proxy for the familiarity of the raters with the co-occurrence of verb and construction; it is also the measure that would register the pre-emptive effect of the infinitive complement construction on the *that*-complement construction (Goldberg 2011; Boyd and Goldberg 2011). In the majority of cases (84/95) the infinitive complement clause is the more frequently used option. For 5/95 verbs, the *that*-complement construction had a higher raw frequency than the infinitival complement construction, and in 6/95 cases both constructions occurred in roughly equal proportions.

Second, the overall frequency of any of the 95 verbs in the *that*-construction was recorded. For each verb, the final search results were manually checked and an estimated number of

[6] The fact that the *that*-clause occurs within sentence boundaries in a language that does not impose a strict word order implies that the contexts potentially run over longer sequences than the 2, 3 or 4-grams for which frequency effects and forward transitional probabilities have been reported in the literature (for a recent overview see Snider & Arnon 2012).

occurrences was calculated on the basis of the percentage of correct hits within the first 200 extractions (as suggested by Amir Zeldes, p.c.). For 26/95 verbs, the *that*-construction was not attested in the 1.5 billion word corpus; since reliable statistical inference is in principle impossible even for hapax (and dis) legomena (Evert 2005: 133, 166) these verbs are removed from the sample and analyzed separately (See Section 3.2.2; a list of these verbs is provided in Appendix 1). Of the attested 69 verbs, 51 occur fewer than 1000 times in the *that*-construction in the 1.5 billion corpus, i.e. fewer than 0.66 times per million words, making the *that*-alternative a "legal" (in the terminology of Caldwell-Harris et al 2012) option at best. This study thus extends the range of frequencies studied further downwards (compare Bannard and Matthews 2008, Snider and Arnon 2012). Of the remaining verbs, 15/95 verbs occur up to 10,000 times or 6.66 ipm in the *that*-construction, and for 3/95 verbs the *that* construction is encountered between 10,000 and 15,000 times, or up to 10 ipm, which is still within the upper bound for low frequency according to the criteria used in Arnon and Snider (2012).

2.2.1.3 Relative frequencies

Merely contextualizing frequencies may be insufficient. Jurafsky (1996) advanced the argument that conditional probabilities are a more appropriate metric than frequencies. He showed that a probabilistic model differs from the frequency-based models traditional in psycholinguistics, with true probabilities essential for a cognitive model of sentence processing. More recently, Wiechmann (2008), who surveyed 47 competing variants of association measures and tested them against experimental data from on-line sentence comprehension, found that minimum sensitivity, a conditional probability, outperforms any of the other measures in predicting reading behavior collected in an eye-tracking experiment.

Probabilities can be conditioned on more than one type of information, an issue that hitherto has received little attention but has caused problems (see Krajewski et al. 2011). Therefore, two types of directional relative frequencies were computed, Attraction and Reliance (Schmid 2000: 56), for both the infinitival and *that*-complement constructions. Relative frequencies do not merely state how many instances of a verb or of a verb*construction combination are found in the corpus, but relate these numbers to, in this case, the total number of verbs or constructions in the corpus. The first unidirectional relative frequency, Attraction, reveals the degree to which a lexico-grammatical pattern attracts a verb, relative to verbs competing on the paradigmatic axis. Attraction is thus the frequency of a verb*construction combination given the frequency of the *that*-construction. The second unidirectional relative frequency measure, Reliance, measures the degree to which a verb depends on a lexico-grammatical pattern, relative to occurrence of the same verb in other patterns. Reliance is defined as the frequency of a verb*construction combination given the frequency of the verb. It thus gives an idea of how likely the construction is to follow if the verb is known. The rank list for Reliance is often topped by lexemes which are highly specialized for occurrence in the given pattern but may be fairly infrequent overall (Schmid 2010: 110). The way in which these relative frequencies are calculated reveals that they are in fact conditional probabilities, although they are not labeled or discussed as such in Schmid (2000, 2010): Attraction equals $p(v|c) = p(v \text{ and } c)/p(c)$, while Reliance is $p(c|v) = p(v \text{ and } c)/p(v)$ (see Gries et al 2005: 660 for a similar conclusion regarding Reliance, which they term faith).[7]

Apart from these two measures that originate within usage-based corpus linguistics, a wide range of association measures are available from within computational linguistics (see Evert 2005 or http://www.collocations.de/AM/contents.html for formulae and explanation). The following were computed (see Appendix 2)[8]: DeltaP, logarithm of the odds ratio, a discounted version of the

7 Schmid (2010: 108) explains that scores for reliance are not proportional to their frequency of occurrence in the construction since the denominator of the fraction varies with the overall frequency of a verb in the corpus.

8 There is also a range of contingency-based measures available that rely on null-hypothesis significance testing to establish the strength of the association, e.g. logarithm of the p-value of the Fischer exact test, logarithm of the p-value of the binomial test, logarithm of the p-value of the Poisson test, z-test score, Student's t-test score,

log odds ratio, logarithm of the relative risk, Pointwise Mutual Information, squared PMI, cubed PMI, local mutual information, log-likelihood ratio, Dice coefficient, Jacard coefficient, minimum sensitivity, geometric mean, logarithm of Poisson likelihood, Poisson-Stirling, logarithm of the hypergeometric likelihood. Significant associations were found between measures of the same type, such as Reliance, Δp, log-odds ratio, log-relative risk and PMI (cf. Levshina (ms.)). This was taken into account while building the model (see Section 3.1).[9]

2.2.2 Related words: morphological family size and morphological transparency

Apart from information on frequency of (co-)occurrence, structural information was included, in particular information on morphological family size and morphological transparency. Morphological family size is a measure of the frequency of inflectionally and derivationally related words; Nagy et al. (1989) were the first to report faster and more accurate processing for words with larger morphological family sizes. Morphological transparency is a measure of the ease with which the 95 verbs could be linked to other words. The motivation for including transparency is that speakers may be able to rate a rare and hence likely unknown verb consistently if they can relate it to a better known word from which they could borrow hints about the word's meaning or its constructional possibilities. These two measures are interlinked: relating verbs to other words is facilitated if the words share an easily identifiable root, and this in turn is sensitive to the relative

Pearson's chi-squared statistic. These are either likelihood measures (which are based on the ratio between the maximum likelihood of the observed data under H_0 and its unconstrained maximum likelihood without making any assumptions about the population parameters), or measures based on (standard) exact and asymptotic hypothesis tests. It has been argued that measures with the capacity to relate observed to expected frequencies, would be superior to those who do not include this information (Schmid 2010: 111-115, Gries 2012), if only from a mathematical point of view. Yet significance tests were designed for small numbers of observations, hence they cannot legitimately be used for our purposes: given the large number of observations available, even very small deviations from independence would yield statistically significant results. This would, however, merely be significant evidence of an insignificant association as p-values measure strength of evidence, not evidence of strength.

9 Attraction correlated much more strongly with frequency than Reliance did. This is mentioned in Schmid (2010:1087): since the denominator of the fraction is the same for all verbs which occur in a pattern, the scores for this value are directly proportional to the raw frequencies of the verbs.

frequency of the derived form and the base (Hay 2001, but see Hanique and Ernestus 2012 for a critique). The expectation would therefore be that a low frequency verb may achieve a high rating if it is transparent and in particular if it is easily related to or derived from a high frequency word or words.

Family size and transparency were calculated in the following manner. For each verb, all related words were culled from word-formation dictionaries for verbs, nouns and adjectives. For all of these derivationally related words, the raw frequency of each word in all its inflectional forms in the IPI-PAN sample corpus was determined, and the transparency of the relation to the verb was quantified. This was done by assigning a score according to the bottom-up constructed procedure outlined in Table (1). The procedure takes into account the nature and number of procedures required to get from the verb to the related word. The label "additions" handles affixations including reflexives while "change" groups vowel and consonant changes to the stem; "change" makes similarities harder to detect and is therefore penalized in the calculation by equating the effect of one change to that of two additions.

The related word belongs to the/a	**score**
same word class with 1 addition or other word class with no further addition or change	5
same word class with two additions or one change or different word class with one addition	4
same word class with one addition and one change or with two changes or with three additions or different word class with two additions or one change	3
different word class with three additions or with two changes or with one addition and one change	2
different word class with two additions and one change	1

Table (1): procedure for assigning a morphological transparency score

For each verb, three related scores were then calculated, i.e. the total number of related words available, the summed raw occurrence frequencies of all related words (it so-called base frequency) and an average transparency score which is the sum of the individual transparency scores of all related words divided by the number of related words. All three scores were considered for inclusion in the regression model.

2.2.3 Semantics

As mentioned (Bondaruk 2004: 208), there is no obvious, i.e. specific, semantic motivation that would unite the 95 verbs into sub-categories that share the (im)possibility of accepting a *that*-clause. Neither do the verbs naturally form smaller subgroups that could be motivated by reference to verbal semantics. Therefore, more abstract semantic properties were looked at. It has been argued that event objectification and event separability underlie a verb's ability to introduce a *that*-clause (Givón 2001, Divjak 2007). A *that*-clause would be more likely after those verbs that treat the infinitive following them as if it were their direct object, as illustrated in example (1'). In (4') the infinitive cannot be treated as the finite verb's object:

(1')

Co	on	zdecydował?	Wyjechać.
What	he	decide PF.IND.PAST.MASC.3SG	Leave PF.INF

What did he decide? To leave.

(4')

Co Marek	ci dał?	*Poprowadzić samochód.
What Marek you give IMPF.IND.PAST.MASC.3SG?		Drive PF.INF car

What did Marek give you? *To drive the car.

A *that*-clause is also more likely to be allowed if both events can be separated in time, as illustrated in examples (1'') and (2''):

(1'') Wczoraj zdecydował jutro wyjechać.

Yesterday he decided to leave tomorrow.

(2'') Wczoraj Marek mi dał poprowadzić samochód jutro.

*Yesterday Marek let me drive the car tomorrow.

Data on the acceptability of event objectification and event separability were collected in the same way as data on the acceptability of *that*-clauses (see Section 2.2 above) from the same respondents. Average event objectification and event separability scores were calculated for the analysis.

3. Results

The distribution of modes for all 95 verbs included in the study is presented in Table 2, and shows that there is a clear acceptability preference in 75/95 cases:

<table>
<tr><th></th><th>-2</th><th>-1</th><th>0</th><th>1</th><th>2</th></tr>
<tr><td>One mode</td><td>26</td><td>12</td><td>14</td><td>10</td><td>13</td></tr>
<tr><td rowspan="4">Multiple modes (adjacent)</td><td colspan="2">4</td><td></td><td></td><td></td></tr>
<tr><td></td><td colspan="2">1</td><td></td><td></td></tr>
<tr><td></td><td></td><td colspan="2">3</td><td></td></tr>
<tr><td></td><td></td><td></td><td colspan="2">2</td></tr>
<tr><td>Multiple modes (non-adjacent)</td><td colspan="5">9</td></tr>
<tr><td>No mode</td><td colspan="5">1</td></tr>
</table>

Table (2): Distribution of ratings by mode

There are 10 verbs that show multiple but adjacent modes while a further 10 show multiple but non-adjacent modes. That is a substantial amount of divergence to account for if we assume a grammar that projects beyond the input data, yet an impressive amount of convergence for a pattern that has to be abstracted from usage.

Of the 10 verbs with multiple but non-adjacent modes, 5 are attested with the *that*-clause in the corpus, i.e. *bać się* (‘be afraid of, fear’), *bronić* (‘defend, guard, vindicate, assert’), *obawiać się* (‘fear, be afraid, be anxious’), *ofiaro(wy)wać się* (no translation available) and *(za)wahać się* (‘hesitate, weaver’). For *bać się* (‘be afraid of, fear’) and *obawiać się* (‘fear, be afraid, be anxious’) it has been noted that a non-default “no obligatory control” or non-coreferential interpretation would be available in combination with a *that*-clause (Bondaruk 2004: 203); it is possible that some raters were familiar with this interpretation and rated the combination therefore as acceptable while others rejected it. The remaining 4 verbs showing conflicting modes are not attested with the *that*-clause in the corpus, I.e. *brzydzić się* (‘abhor, loathe, have an aversion’), *krępować się* (‘be embarrassed, feel uneasy’), *potrafić* ('know how to, manage') and *śmieć* (‘dare, venture'). One further verb (*godzić się* 'agree, consent') that is attested with the *that*-clause in the corpus has all acceptability levels as mode.

3.1 Model fitting and performance

Both linear and ordinal regression models were fitted in a step-wise forward fashion to determine which of the variables considered has the largest impact on the acceptability ratings, how the variables relate to each other and how much of the variation they explain. Given that linear models did not handle the extremes of the distribution well and hence yielded a lower accuracy rate for the classification results, an ordinal model was preferred. Two sets of ordinal models were run, with and without control variables, using the generalized linear mixed models procedure for ordinal data in SPSS 21.

The resultant minimally most adequate mixed model for the 1031 ratings pertaining to the 69 verbs that are attested in the *that*-construction in the corpus contains a random effect for item and one for rater to account for the fact that each item had been rated 15 times and each rater had seen five experimental sentences. With the random effects in the model, the linguistic variables that individually make a significant contribution are, in order of coefficient size and strength of evidence:

i) pattern transparency (coeff=.198, st.err.=.062, p=.002),

ii) one of the following four highly correlated association measures: the logarithm of the Reliance of a verb on a construction (coeff=.143, st.err.=.051, p=.005), the logarithm of the odds ratio (coeff=.142, st.err.=.050, p=.005), the relative risk for the verb (coeff=.142, st.err.=.050, p=.005) or pointwise mutual information (coeff=.143, st.err.=.051, p=.005).[10]

[10] Evert & Krenn (2001) found that Mutual Information systematically overestimates the collocativity of low-frequency pairs but low frequency was defined as $2 \leq f < 5$ and in this sample there are only 2 verbs that would fall

iii) the average event separability score (coeff=.416, st.err.=.152, p=.006)

iv) the logged (estimated) frequency of the verb in the *that*-construction (coeff=.110, st.err.=.049, p=.025).

None of the other variables discussed in Section 2 improves the base model with two random effects significantly. Unigram frequencies for the verb and the joint or relative frequency with which the verb occurs with the alternative and potentially pre-empting infinitival construction appear to play no role in predicting acceptability of the verb in the *that*-construction; neither do abstract constructional semantics improve prediction accuracy.

Adding pattern transparency and an association measure to the model simultaneously (pattern transparency coeff=.178, st.err.=.060, p=.003; logarithm of the Reliance of a verb on a construction coeff=.124, st.err.=.049, p=.011) annuls the initially significant effect of average event separability score and (estimated) frequency of the verb in the *that*-construction. Variable interactions and by-subject random slopes for the fixed variables of interest did not significantly improve model fit.

A second set of models was run with fixed effects including control variables to ensure that the observed effects of the linguistic variables transparency or frequency should not be attributed to known properties of subject or materials. Of all control variables considered, only rater generosity appeared to make a significant contribution (coeff=1.459, st.err=.151, p=.000); position of the experimental item in the experiment did not have a significant effect (coeff=.071, p=.081). The resulting models were virtually identical to the model without control variables, except for a small change in the coefficient estimates and associated p-values of the retained explanatory variables (pattern transparency coeff=.184, st.err.=.062, p=.003; logarithm of the Reliance of a verb on a

in this range.

construction coeff=.117, st.err.=.050, p=.02). Transparency scores were also considered binned[11], which showed that the lowest bin decreases ratings most (bin 1 coeff=-.752, st.err.=.287, p=.009; bin 3 coeff=-.0482, st.err.=.367, p=.19; bin 4 coeff=-.117, st.err.=.286, p=.682).

Finally, a third set of models was run to establish whether Transparency and Reliance have the same effect across the verb frequency spectrum. This was achieved by splitting Transparency and Reliance up by raw verb frequency quartiles; the quartiles were entered as levels of an ordered factor. Transparency correlates significantly with Acceptability across verb frequency quartiles (Q1 coeff=.223, st.err.=.097, p=.022; Q2 coeff=.158, st.err.=.078, p=.044; Q3 coeff=.202, st.err.=.081, p=.013; Q4 coeff=.172, st.err.=.084, p=.040). In the Transparency-by-verb-frequency quartile model, the Reliance estimate changes slightly (coeff=.113, st.err.=.054, p=.035). In the Reliance-by-verb-frequency-quartile model the second quartile makes a significant contribution (coeff=.142, st.err.=.066, p=.033) while the third quartile comes out borderline (coeff=.112, st.err.=.057, p=.052) and the fourth is likely not significant (coeff=.104, st.err.=.058, p=.073). In the Reliance-by-verb-frequency-quartile model, the pattern transparency estimate changes slightly (coeff=.194, st.err.=.068, p=.005).

Because information is lost when numerical variables are factorized, the prediction accuracy results of the second stage are reported. This model predicts the actual rating assigned by the subjects relatively accurately. Overall, the model predicts the rating correctly in 43.4% of cases, with the highest values on the diagonal for each rating and the second highest value in an adjacent cell (see Table 3). This is nearly twice as good as achieved by randomly assigning ratings for a 5-way choice. Taking into account that corpora do not match experience with language perfectly, and certainly not every individual's experience with language, this is a respectable result for an ordered

---

11 Transparency scores were recoded into 6 categories of equal width using cut-off points of 0, 5/6, 2*5/6 and so on; groups that were giving very similar coefficients were combined, yielding 5 categories, 0 to 0.833, 0.833 to 1.67, 1.67 to 2.5, 2.5 to 3.75 and 3.75 to 5; the second bin contained only 15 elements and was not estimated; the coefficient for the fifth bin was set to zero as it appeared redundant.

5-way choice. The figure rises to 63.4% if we turn the 5-way choice into a 3-way choice by collapsing [-2 and -1] and [1 and 2] and to 83.8% if we allow the model to be 1 point off in either direction.

	Predicted				
Observed	**-2**	**-1**	**0**	**1**	**2**
-2	84	58	34	10	1
-1	42	83	47	31	1
0	11	40	83	72	14
1	2	22	54	112	37
2	2	9	27	70	85

Table (3): Classification table for a mixed effects model with fillers, transparency and reliance as predictors

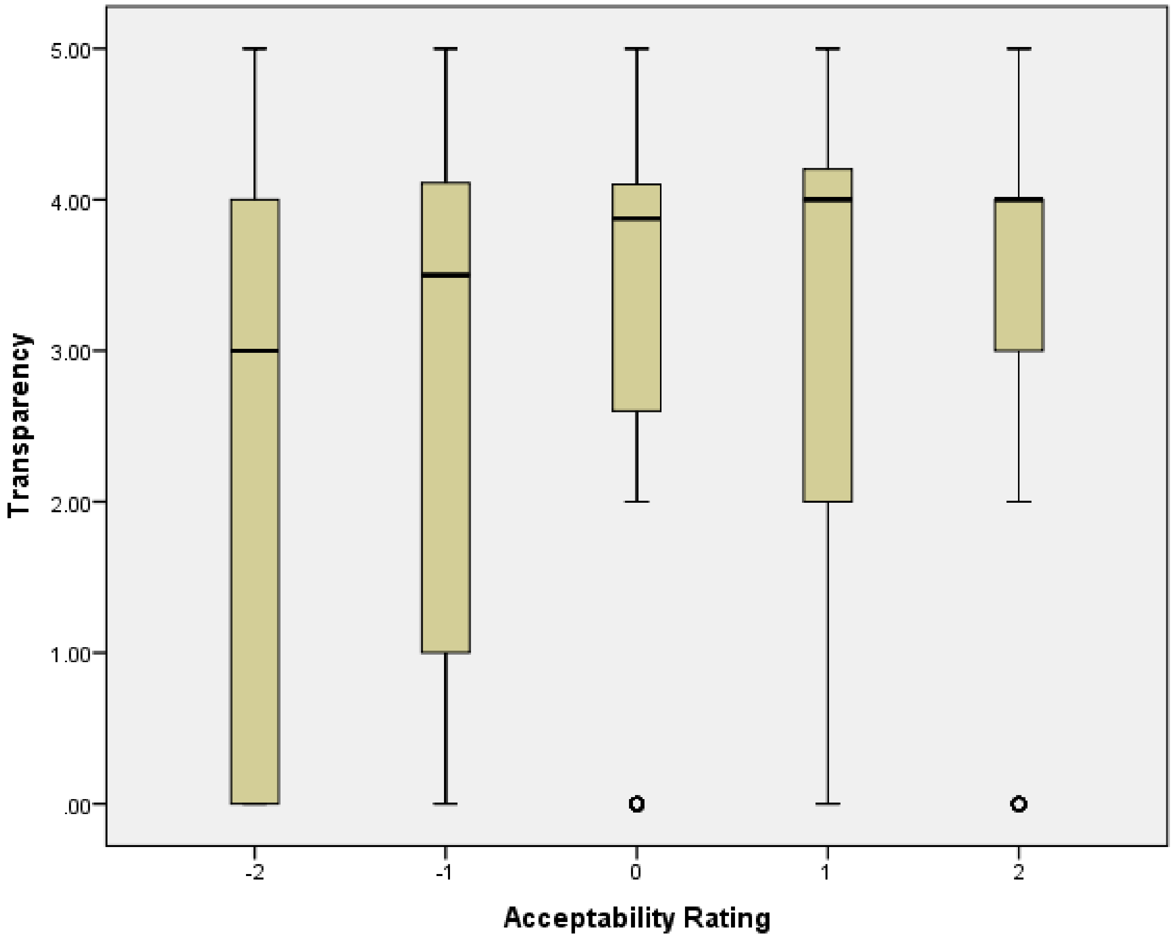
3.2 Interpretation of the results

Overall, the single strongest predictor for experimental sentence acceptability appears to be the rating given by the raters to the filler items (coeff=1.469, st.err=.152, p=.000). With estimates for this predictor going up as the experimental sentence scores go up, subjects who give fillers higher ratings are more likely to give experimental sentences higher ratings too. This predictor thus captures (some of) the non-linguistic variability associated with the acceptability rating procedure, i.e. the fact that some raters are more generous than others.

3.2.1 Analysis of results for attested lexico-syntactic combinations

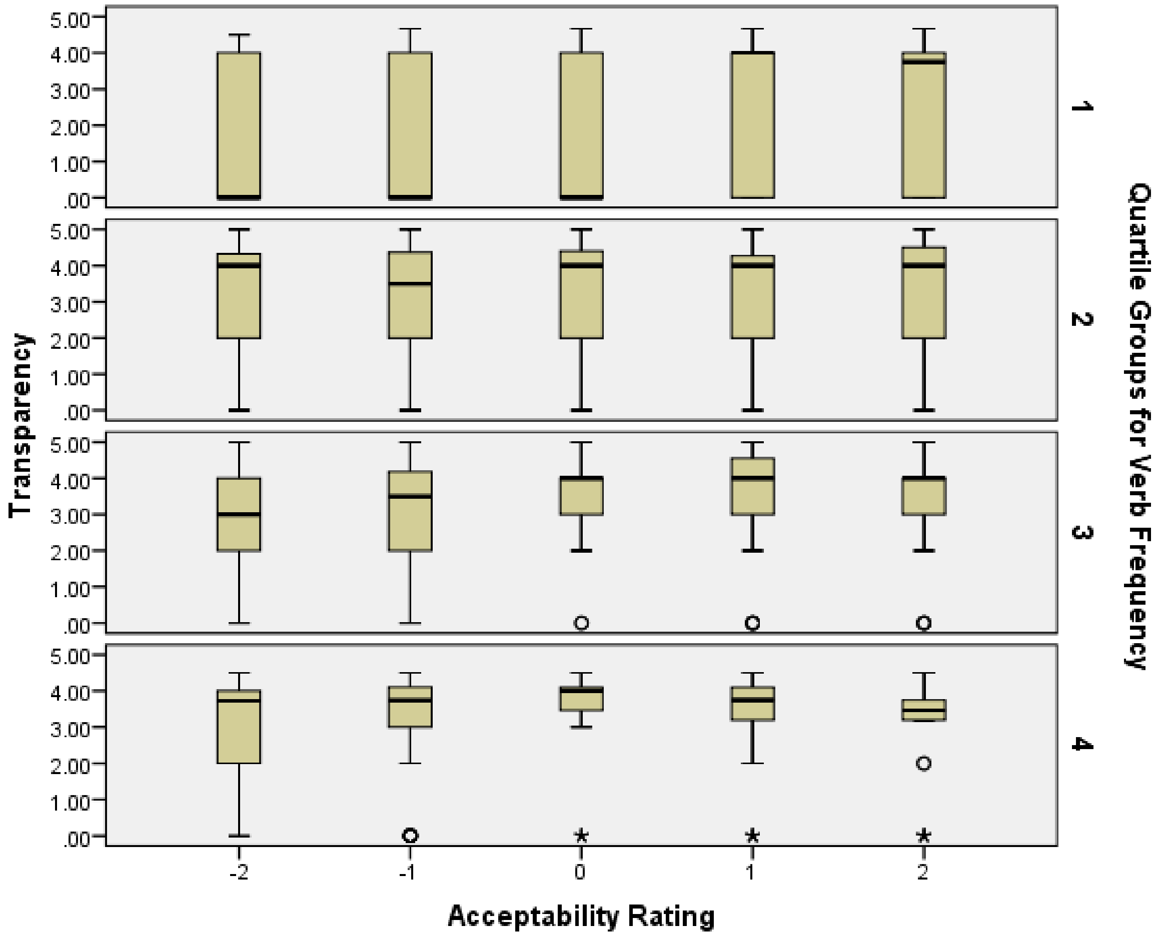
A first significant contributor to a model of the acceptability of a verb in the *that*-construction is the average morphological transparency of that verb: the more recognizably the verb is related to other words, the higher the acceptability rating. As Figure (1) shows, the transparency score on the Y axis increases with each score-point on the X axis. While lexico-syntactic combinations receiving low ratings show a wide range of transparency scores, the height of the range reduces as the rating goes up, and is the shortest for the highest rating.

Figure (1): Box and whiskers plots of average transparency score by acceptability rating



The effect of transparency on acceptability is shown split up per quartile of verb frequency in Figure (2). Transparency is a significant predictor of acceptability for all quartiles of verb frequency. As mentioned in Section 3.1., binning transparency shows that the effect is strongest for non-transparent patterns: it is lack of transparency that lowers the ratings significantly. The total number of words a verb is related to, i.e. the size of the morphological family, or the morphological family frequency, i.e. the summed frequencies of all related words, do not significantly improve the model.

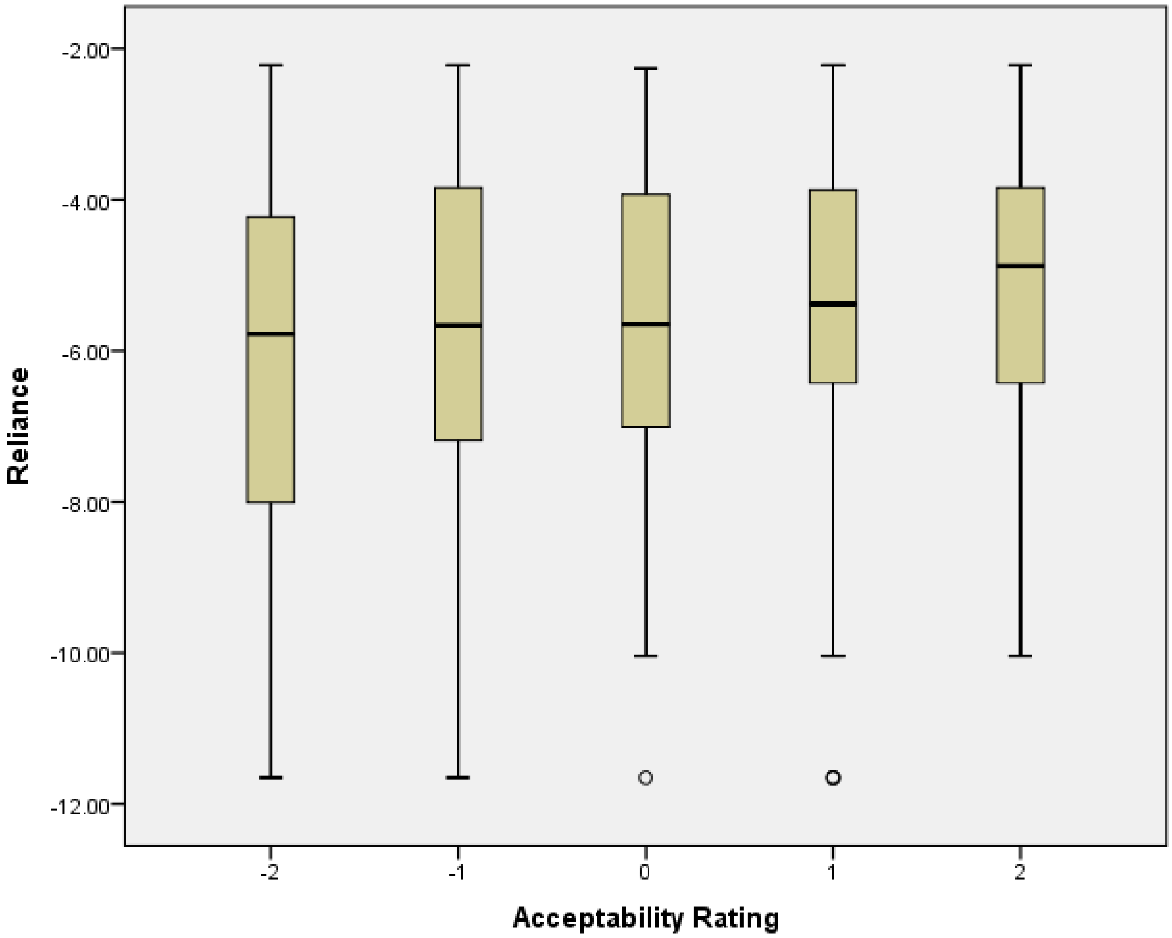
Figure (2): Box and whiskers plots of transparency score by acceptability rating per quartile of verb occurrence (0.4 - 6; 10 - 41; 50-140; 188 - 1516 ipm



Being clearly part of a morphological family is thus more important than the exact size of that family or the precise number of times members of the family are encountered. Although there is work within morphology suggesting that the family size effect is a semantic effect with a morpho-syntactic component (De Jong, Schreuder, Baayen 2000: 359) and that the morphological context in which a word appears influences the way in which activation spreads in the mental lexicon (Bertram, Baayen and Schreuder 2000), further research is needed to unravel the mechanisms of transparency and its effects on syntactic acceptability. The effect suggests that verbs that are visibly part of a morphological family are (considered) constructionally more versatile: speakers appear to assume a wider variety of constructional options for words that are part of morphological families. There are at least two possible explanations for this. It is plausible that the related words directly lend their constructional possibilities to the verb: if there is a relative that allows the *that*-clause, this option is more likely to be allowed for the verb tested. It is also possible that related words help access (more of) the meaning of the verb tested, and if that meaning is deemed compatible with the meaning of that *that*-clause, this clause is rated more highly.

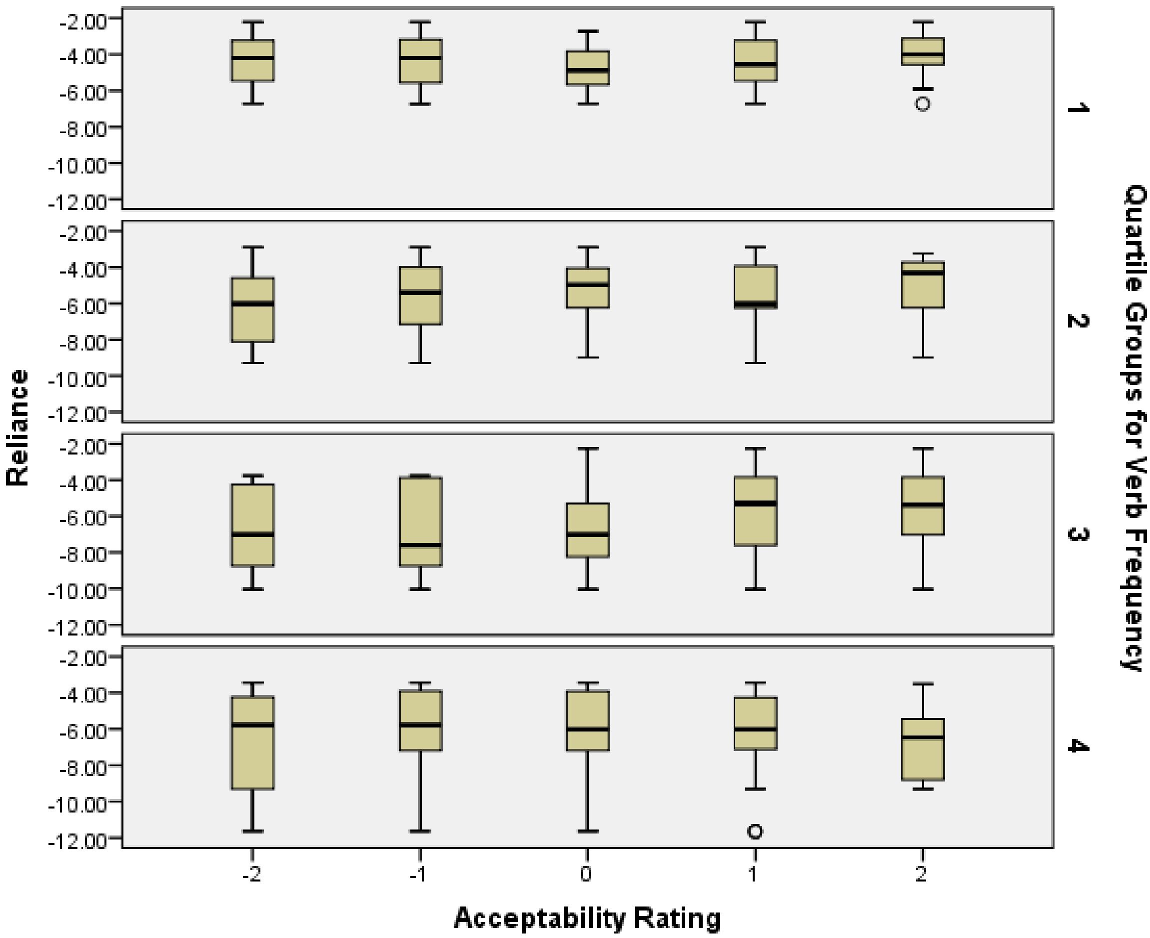
A second significant linguistic predictor is the (logarithm of the) Reliance of the verb on the construction, or any of the association measures that Reliance is strongly correlated with: as Reliance or any of its correlated association measures goes up, so does the rating, and this occurs across the range of acceptability ratings assigned to the lexico-constructional combination (see Figure 3). In addition to a small but significant increase in the median, the lower bound also increases stepwise. Because Reliance is the simplest measure that makes no assumptions about the data, I will focus on Reliance in the remainder of this paper.

Figure (3): Box and whiskers plot of the (logarithm of the) Reliance of the verb on the construction by acceptability rating



Plotting the data split up into quartiles according to the raw frequency of occurrence of the verb shows how the second and third quartiles of raw verb frequency drive this effect, that is verbs with raw frequencies between 10 and 140 instances per million words. There is a weaker positive relationship for the top 25% of the data (with frequencies between 188 and 1516 ipm), and an insignificant positive relationship for the bottom 25% (with frequencies between 0.4 and 6 ipm).

Figure (4): Box and whiskers plot of the (logarithm of the) Reliance of the verb on the construction by acceptability rating per quartile of verb occurrence (0.4 - 6; 10 - 41; 50-140 ; 188 - 1516 ipm)



The plot in Figure (4) reveals further that, in the first quartile, the relation between rating and Reliance is U-shaped, with ratings of 0 assigned to those combinations that have on average the lowest Reliance value, as calculated on the basis of the corpus data. The second and third quartiles show the expected linear effect, with higher Reliance values by and large yielding higher ratings; the second quartile shows a steady increase in average Reliance value except for items that received rating 1, while in the third quartile there is a dip in average Reliance score for items rated as -1. In the fourth quartile the average Reliance value decreases slightly as the rating goes up. More precisely, in the 4th quartile, the median and range of Reliance values are broadly similar across ratings, except for combinations assigned a rating of -2, for which the range spans -4 to -9 rather than -5 to -9. Furthermore, combinations assigned a 2 stem from a Reliance range that is positioned lower, i.e. while the range for ratings -2, -1 and 0 runs roughly between -4 and -7, the range for rating 2 spans -5 to -9.

3.2.2 Analysis of non-attested lexico-syntactic combinations

Verbs that are not attested in the *that*-clause in the 1.5 billion word corpus were analysed separately using rank order correlations. For these combinations, the median acceptability rating is -1, except for verbs in the bottom quartile for raw verb frequency where it is 0; slightly higher ratings are thus assigned to verbs in *that*-clauses if those verbs are of lower frequency (ρ = -.258, p= .000). Likewise, higher ratings are assigned to verbs that are less transparent (ρ =-.178, p = . 000). As far as these unattested combinations of verbs and *that*-clauses are concerned, the tendencies are thus reversed: speakers appear more likely to accept alternative constructions for verbs that they do not encounter often and for verbs that they cannot easily relate to other words. Ambridge (2013) reports a similar effect for children who seem more accepting of low frequency verbs being used in novel high frequency constructions, than of high frequency verbs being used in alternative constructions. In this particular case, the effect is in large part due to a number of very high frequency verbs such as modal verbs that are part of this category: modal verbs occur exclusively with an infinitival clause and categorically reject the *that*-clause. This shows that for variation to be pre-empted in the adult syntactic system, a very strong preference for one constructional alternative appears to be required.

4. Discussion

This study has quantified the effect that frequency of occurrence has on acceptability for low-frequent phenomena at the syntax-lexis interface: two thirds of verbs studied occur fewer than 0.66 ipm with the *that*-clause, and this extends the range of frequencies included in this study below the lower bound set in previous research in this area (cf. Caldwell-Harris et al. 2012). The reported results were obtained in an off-line acceptability rating experiment that reduces real-time processing pressures.

Rater generosity explains the lion-share of the variance in ratings: a subject's general tendency to assign low, medium or high ratings is highly predictive for the acceptability ratings assigned to individual items. Linguists should thus be acutely aware of the fact that (linguistic acceptability) judgments reflect properties of the rater, rather than properties of the grammar. Analyses of linguistic data that do not take this variable into account are very likely to overestimate the explanatory power of the linguistic variable(s) of interest (cf. Rohde 2003; Sprouse 2013). This finding is also interesting from the point of view of language change. The low frequency lexico-syntactic combinations rated in the surveys occupy a position at the fringes of the language system. It may well be thanks to generous raters, who are more accepting of rare combinations and assign high ratings, that verbs start being used in novel constructional frames. And by using them, their overall frequency increases, allowing the structure to gain ground. Further research will reveal whether rater generosity is one of the relevant predictors for and hence driving forces behind language change and whether low frequency combinations could embody the potential of a language for change and expansion (cf. Bybee: 2010: 101).

Contrary to what has been assumed in the literature on frequency and acceptability, the effect of frequency does carry through to the acceptability of syntactic structures - often considered a rule-governed phenomenon par excellence – an this goes against the claim that usage would not affect grammar. The effect is attested for combinations most of which are encountered fewer than 0.66 times per million words (contra Alegre and Gordon 1999 but see Baayen et al. 1997): in a model that includes subject and item as random effects, an increase in the probability of encountering a verb in a *that*-clause given the frequency of the verb results in an increase in acceptability rating. Counts (in particular uncleaned) of the frequency with which verbs occur in *that*-clauses do not contribute significantly (cf. Baayen 2010; McDonald and Shillcock 2001; Raymond and Brown 2012; Recchia, Johns and Jones 2008; Theakston 2004: 28), neither when entered as sole factor, nor when entered into the model after Reliance; in the latter case, co-occurrence frequency shows a small negative but insignificant effect. The results thus confirm that frequency information forms an integral part of a speaker's linguistic knowledge (or is a very good proxy for a range of other things, see Baayen 2010), while at the same time stressing that, for understanding lexico-syntactic acceptability data, frequency of occurrence is best modeled using conditional probabilities. The fact that raw co-occurrence or n-gram frequency is not predictive whereas conditional probabilities are predictive, even for phenomena that occur less than 0.66 times per million words, suggests that these combinations are not evaluated on the basis of "raw exemplars" but rather on the basis of configurations of internally structured exemplars. In other words, the observed Reliance effect is not a result of simply hearing the verb with any occurrence of *that* but stems from higher-order knowledge of the existence of *that*-clauses that linguists describe as schemata or rules (cf. Wiechmann 2008). Existing successful n-gram models may well have shown reasonable performance because the phenomena modeled happened to involve adjacent words in languages with fixed word order like English. Furthermore, a lower and upper word frequency threshold were identified where an increase in conditional probabilities starts or stops causing an increase in acceptability of a verb in a *that*-clause (cf. Erker and Guy 2012): it is the middle 50% of the data (i.e. the middle quartiles of verb unigram frequency) that shows the effects of Reliance most strongly. Speakers seem to entertain another strategy for the lowest frequency verbs (i.e. verb in the bottom quartile of verb unigram frequency), and they rely on information gleaned from pattern transparency for guidance. The highest frequency verbs are likewise more resistant: they are possibly so well-engrained that properties in addition to co-occurrence need to be respected (cf. Bybee 2006: 715).

On a methodological level, these findings show that it is not so much the case that usage frequency has problems predicting acceptability judgments at the low end of the frequency spectrum. It is rather the case that the wrong type of frequency data has been foregrounded: focus has been on raw or contextual/co-occurrence frequency or on measures that relate observed to expected frequencies rather than on higher-dimensional conditional probabilities. For the prediction of the acceptability of unusual lexico-syntacticcombinations, the simplest measure, Reliance, a uni-directional measure without corrections for expected frequency, performed better than or equally well as a range of mathematically more sophisticated measures. Reliance of the construction on the verb captures a conditional probability which in its logarithmic form is the negative of Surprisal. Surprisal or self-information, an information-theoretical concept (Cover and Thomas 2006), has been gaining ground in psycholinguistic models of syntactic processing (for a first implementation see Hale 2001, followed by Levy 2008 and Fernandez Monsalve et al. 2012) and fMRI evidence of its role in comprehension has been described (Willems et al 2015). Surprisal is related to the surprise ratio (Barlow 1990) and variants thereof that have been implemented in computational emergentist models, e.g. ADIOS in Solan et al. (2005). Although Surprisal is conventionally calculated using a logarithm base 2 rather than the natural logarithm used here, it is the negative logarithm of a probability, i.e. the negative of the logarithm of the conditional probabilities used in this paper. Hale (2001) argued that the difficulty of a word should be proportional to its Surprisal (its negative log-probability) in the context within which it appears.

High Reliance of a verb on a construction thus implies low surprise to encounter this construction after that verb, which increases the rating for the construction: the less surprised raters are to encounter a *that*-clause after a verb that is also (and in the majority of cases more frequently) followed by an infinitive, the higher the rating for the *that*-paraphrase. In adult speakers, lexico-syntactic variation trumps pre-emption (cf. Braine and Brooks (1995:368) who describe that with age there appears to be an increase in flexibility in switching between sentence constructions to meet conversational demands).[12]

The findings also cast further doubt (see also Schmid and Küchenhoff 2013/2015) on the primacy of and need for mathematically superior formulae relying on expected frequencies, at least for off-line structure rating activities. This makes conditional probabilities the psychologically more relevant variable, which is in line with what has long been known about context-dependent memory (Greenspoon and Ranyard 1957; Godden and Baddeley 1975), the role of prediction in cognition (Bar 2009) and the extraction of reliable and non-redundant information that correlates with reinforcement as core ingredient of learning (Rescorla and Wagner 1972; Sutton and Barto 1981).

Because speakers cannot know a priori whether a verb will prefer one syntactic context or freely alternate between a few, it is very likely that co-occurrence information of all frequencies is registered. Different from Keller (2003), who argued that data sparseness makes it implausible that the human parser directly records structural frequencies, I would argue that the fact that low frequencies cause problems making predictions about the acceptability of a structure does not provide evidence to the contrary. Instead, this situation is expected on a probabilistic approach as prediction mechanisms cannot make reliable inferences about the acceptability of a structure

[12] When entered into the model, the logarithm of the reliance of a verb on the infinitival construction gave a positive but non-significant (.2) coefficient. This would mean that the effect of the two types of reliance would be additive, and indicate that to speakers it is more important that the verbs belong together in a construction, than how they are joined up.

without sufficient data, i.e. without knowledge of its components. This difficulty can also be explained by what we know from memory research, in particular from research on how information is transferred from immediate working memory to long term memory (Gupta 2012). Learning human languages draws on both the procedural and declarative memory systems. Novel mappings require creating new pathways between inputs and outputs, and thus may be initially stored as part of episodic memory. If that novel information is never encountered again, the weighted connections that represent it will be overwritten as new patterns are encountered. But if that stimulus is repeatedly encountered, each exposure provides another training trial in which it can be integrated into long-term memory structures (Divjak and Caldwell-Harris 2015: 64). Conditional probabilities suffer from data sparseness too, since smaller samples are less representative than larger ones. And the higher the level of uncertainty surrounding a prediction, the smaller the likelihood should be of receiving consistent ratings. To counter this, for the rarest verbs, speakers may well entertain another strategy and rely on information gleaned from pattern transparency for guidance, as suggested by this model.

5. Conclusions

A number of studies in both the generative and usage-based traditions had reported that corpus frequencies are poor predictors for off-line acceptability ratings in morphology and syntax, in particular at the lower end of the frequency spectrum. Because judgments of grammaticality provide a substantial part of the empirical foundation of many areas of linguistics and many linguistic theories, it is crucial to gain a better understanding of how these judgments relate to one of the core concepts of contemporary linguistic theory: frequency of occurrence. This study used data from a 1.5 billion word corpus of Polish to investigate the factors that predict the acceptability of a verb in a *that*-clause when the occurrence of the *that*-clause is low frequent (<0.66 ipm), highly variable and not motivated semantically.

A mixed effects ordinal regression model showed that a significant linguistic predictor is the average morphological transparency of the verb: the more recognizably the verb is related to other words, the higher the acceptability rating. This is true in particular for verbs that are in the bottom quartile (0.4 to 6 ipm) for unigram frequency. Probabilities of encountering a verb in that *that*-clause conditioned on the verb explain a further significant share of the variance in subjects' rating behaviour: an increase in the degree to which a verb relies on that *that*-construction in the corpus correlates positively with an increase in the acceptability of the verb in the *that*-clause. This effect is most strongly observed for verbs with unigram frequencies in the middle quartiles (10 to 140 ipm). In other words, a lower and upper word frequency threshold were identified where an increase in conditional probabilities starts or stops causing an increase in acceptability of a verb in a *that*-clause. Finally, the results suggest that acceptability judgments are based on configurations of internally structured exemplars, i.e. on higher-order knowledge that linguists describe as constructions or schemata.

These findings show that learning mechanisms tapping into contextual probability distributions are well-equipped to handle the challenge posed by low-frequency items that exhibit subcategorization variation, which has important theoretical consequences. First, the results show that speakers can reliably formulate direct hypotheses about the acceptability of configurations once sufficient evidence about the core component has accumulated. This does not imply that items below this threshold would not leave traces in memory; without such initial traces, the threshold at which frequency becomes a force to be reckoned with would never be reached. Second, the fact that the gradience observed in the acceptability ratings for *that*-constructions reflects the conditional probability of encountering the verb given the construction strengthens the view that implicit probabilistic knowledge is a core component of syntactic knowledge (cf. Bresnan 2007, Bresnan and Ford 2010). Finally, the results capture a crucial step in how grammar emerges from usage by showing that implicit probabilistic syntactic knowledge is based on configurations of internally structured exemplars that accumulate over time. These conclusions not only underscore the importance of fine-grained linguistic data analysis in further work examining the effects of frequency but also contribute to the development of usage-based theories that assign a powerful explanatory role to input frequencies while remaining silent on what needs to be counted and how and further our understanding of how linguistic knowledge fits into human cognition in general.

List of references

Alegre, M. & P. Gordon. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language,* 40, 41-61.

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks.*Bulletin of the Psychonomic Society,* 15(3), 147-149.

Ambridge, B. (2011). Paradigms for assessing children's knowledge of syntax and morphology in E. Hoff (ed.) *Guide to Research Methods in Child Language*. London: Blackwell-Wiley [downloaded from http://pcwww.liv.ac.uk/~ambridge/Downloadable%20Publications.htm on 16 September 2011]

Ambridge, B. (2013). How do children restrict their linguistic generalizations: An un-grammaticality judgment study. *Cognition,* 125: 49–63.

Ambridge, B & E. Lieven. (2011). *Child language acquisition. Contrasting theoretical approaches.* Cambridge: Cambridge University Press.

Arppe, A. & J. Järvikivi. (2007). Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory,* 3(2), 131–159.

Baayen, R. H. (2001). *Word Frequency Distributions*. Kluwer Academic Publishers, Dordrecht.

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

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

Bader, M. & J. Häussler. (2009). Toward a model of grammaticality judgments. *Journal of Linguistics,* 45, 1–58.

Bannard, C. & D. 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.

Bańko, M. (2000). *Inny Słownik języka polskiego PWN*. Warszawa: PWN.

Bar M. (2009). Predictions: a universal principle in the operation of the human brain. Introduction. Philosophical Transactions of the Royal Society B: Biological Sciences, 364:1181–1182

Barlow, H. B. (1990). Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research,* 30, 1561–1571.

Bermel, N. & L. Knittl. (2012a). Morphosyntactic variation and syntactic constructions in Czech nominal declension: corpus frequency and native-speaker judgments. *Russian Linguistics,* 36 (1), 91–119.

Bermel, N. & L. Knittl. (2012b). Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. *Corpus Linguistics and Linguistic Theory.*

Bertram, R., R.H. Baayen & R. Schreuder. (2000). Effects of family size for complex words. *Journal of Memory and Language*, 42, 390-405.

Bod, R., J. Hay & S. Jannedy (eds.). (2003). *Probabilistic Linguistics.* Cambridge, MA: MIT Press.

Bondaruk, Anna. (2004). *Pro and control in English, Irish, and Polish -- a minimalist analysis* .Lublin: Wydawn. KUL.

Boyd J.K. & A. E. Goldberg. (2011). Learning what not to say: the role of statistical preemption and categorization in “a”-adjective production. *Language,* 81(1), 1- 29.

Braine, M. D.S. & P. J. Brooks. 1995. Verb argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: young children's acquisition of verbs* (pp. 352-376). Hillsdale, NJ: Erlbaum.

Bresnan J. (2007). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In S. Featherston & W. Sternefeld (Eds.), *Roots: Linguistics in Search of Its Evidential Base* (pp. 77-96). Berlin: Mouton de Gruyter. [Studies in Generative Grammar].

Brysbaert, M., & Diependaele, K. (2013). Dealing with Zero Word Frequencies: A Review of the Existing Rules of Thumb and a Suggestion for an Evidence-Based Choice. *Behavior Research Methods*, *45*, 422-430.

Bybee, J. (2006). From usage to grammar: the mind's response to repetition. *Language*, 82(4), 711-733.

Bybee, J. & D. Eddington. (2006). A usage-based approach to Spanish verbs of 'becoming. *Language*, 82(2), 323-355.

Bybee, J. & P. Hopper (eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Caldwell-Harris, C., J. Berant & Sh. Edelman. (2012). Measuring Mental Entrenchment of Phrases with Perceptual Identification, Familiarity Ratings, and Corpus Frequency Statistics. In D. Divjak & S. Gries (Eds.) *Frequency effects in language representation* (pp. 165-194). Berlin: Mouton de Gruyter.

Cattell, James M. (1886). The time it takes to see and name objects. *Mind* 41: 63-65.

Chater, N., J. tenenbaum & A. Yuille. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, 10(7), 287 - 291

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Cover, Th. & J. Thomas. (2006). *Elements of Information Theory*. *Second Edition.* Hoboken, NJ: John Wiley.

Cowart, W. (1989). Notes on the Biology of Syntactic Processing. *Journal of Psycholinguistic Research,* 18(1), 89–103.

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments.* Thousand Oaks, CA: Sage Publications.

De Jong, N. H., Schreuder, R. and Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, 15, 329-365.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology,* 25: 108–127.

Divjak, D. (2007). Degrees of Verb Integration. A Binding Scale for [VV] Structures in Russian. In: Divjak, D. & A. Kochańska (eds.) *Cognitive Paths into the Slavic Domain* (pp. 221-256). Mouton de Gruyter: Berlin-New York. [Cognitive Linguistics Research].

Divjak, D. (2008). On (in)frequency and (un)acceptability. In B. Lewandowska-Tomaszczyk (ed.), *Corpus linguistics, computer tools and applications – State of the art* (pp. 213–233). Frankfurt: Peter Lang.

Divjak, D. & St. Th. Gries (eds.). (2012). *Frequency Effects in Language Representation* Volume 2. Berlin: De Gruyter.

Divjak , D. & C. Caldwell-Harris. (2015). Frequency and entrenchment. In E. Dąbrowska & D. Divjak, (eds.), *Handbook of Cognitive Linguistics* (pp. 53 -75). Berlin: De Gruyter-Mouton.

Dziwirek, Katarzyna. (1998). Reduced construction in Universal Grammar: Evidence from Polish Object Control constructions. *Natural Language & Linguistic Theory,* 16, 53-99.

Dziwirek, Katarzyna. (2000). Why Polish doesn't like infinitives. *Journal of Slavic Linguistics*, 8, 57-82.

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

Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition,* 24, 143–188.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics,* 27 (1), 1-24.

Erker D. & G. R. Guy. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language*, 88(3), 526-557.

Evert, S. (2005). *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. Published PhD Dissertation, Institut für maschinelle Sprachverarbeitung, University of Stuttgart. [http://elib.uni-stuttgart.de/opus/volltexte/2005/2371/pdf/Evert2005phd.pdf]

Featherston, S. (2005). The Decathlon model of empirical syntax. In: S. Kepser & M. Reis (eds.), *Linguistic evidence. Empirical, theoretical and computational perspectives* (pp. 187-208). Berlin – New York: Mouton de Gruyter.

Fernandez Monsalve, I., S.L. Frank & G. Vigliocco. (2012). Lexical surprisal as a general predictor of reading time. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 398-408). Avignon, France: Association for Computational Linguistics.

Gilquin, G. & St. Th. Gries. (2009). Corpora and experimental methods: a state-of-the-art review. *Corpus Linguistics and Linguistic Theory,* 5(1), 1-26.

Givón, T. (2001). *Syntax: a Functional Typological Introduction. Vol. 2.* Amsterdam/Philadelphia: John Benjamins.

Godden, D.R., and Baddeley, A.D. (1975). Context-dependent memory in two natural environments: on land and underwater. *British Journal of Psychology*, 66, 325-332.

Goldberg, A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics,* 22(1), 131–154.

Goldberg A E., D. Casenhiser & N. Sethuraman. (2004). Learning Argument Structure Generalizations. *Cognitive Linguistics,* 15(3), 289-316.

Greenspoon, J., and Ranyard, R. (1957). Stimulus conditions and retroactive inhibition. *Journal of Experimental Psychology*, 53, 55-59.

Gries, St. Th. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics,* 13(4), 403-437.

Gries, St. 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* (pp. 197-212). Amsterdam: Rodopi.

Gries, St. Th. (2012). Frequencies, probabilities, association measures in usage-/exemplar-based linguistics: some necessary clarifications. *Studies in Language* 36(3), 477-510.

Gries, St. Th. & D. Divjak (eds.). (2012). *Frequency Effects in Language Learning and Processing* Volume 1. Berlin: De Gruyter.

Gries, St. Th. & A. Stefanowitsch. (2004). Extending collostructional analysis: a corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics,* 9(1), 97-129.

Gries, St. Th., B. Hampe & D. Schönefeld. (2005). Converging evidence: bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics,* 16(4), 635-676.

Grzegorczykowa, R. (2006). *Wykłady z polskiej składni.* Warszawa: Wydawnictwo naukowe PWN.

Gupta, P. (2012). Word learning as the confluence of memory mechanisms: Computational and neural evidence. In M. Faust (Ed.), *Handbook of the Neuropsychology of Language, Volume 1* (pp. 146-163). Oxford: Wiley.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies* (pp. 1–8). Stroudsburg, PA

Hanique, I., & Ernestus, M. (2012). The role of morphology in acoustic reduction. *Lingue e Linguaggio*, 11(2), 147-164

Hasher, L. and R. T. Zacks. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist,* 39: 1372.

Hay, J. (2001). Lexical Frequency in Morphology: Is Everything Relative? *Linguistics: An Interdisciplinary Journal of the Language Sciences,* 39(6), 1041-1070.

Hofmeister, P., I. Arnon, T.F. Jaeger, I.A. Sag & N. Snider. (2012). The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes,* 28(1/2), 1-40.

IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.

IPI-PAN Corpus Sample [2.sample.30.bin.tar.bz2], searchable at http://korpus.pl/; 2nd edition of the IPI PAN Corpus (March 2006) [http://korpus.pl/index.php?page=download]

Janus, D. & A. Przepiórkowski. (2006). POLIQARP 1.0: Some technical aspects of a linguistic search engine for large corpora. In: J. Waliński, Kredens, K. & Goźdź-Roszkowski, S., *The proceedings of {P}ractical {A}pplications of {L}inguistic {C}orpora 2005*. Frankfurt: Peter Lang.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science,* 20, 137–194.

Kaiser, E. (2013). Experimental paradigms in psycholinguistics. In: R. J. Podesva & D. Sharma (eds.), *Research Methods in Linguistics* (pp. 135-168). Cambridge: Cambridge University Press.

Keller, F. (2003). A Probabilistic Parser as a Model of Global Processing Difficulty. In R. Alterman & D. Kirsh, eds., *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 646-651). Boston.

Kempen, G. & K. Harbusch. (2005). The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German clauses. In Stephan

Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical and computational perspectives* (pp. 329–349). Berlin and New York: Mouton de Gruyter.

Kempen, G. & K. Harbusch. (2008). Comparing linguistic judgments and corpus frequencies as windows on grammatical competence: A study of argument linearization in German clauses. In Anita Steube (ed.), *The discourse potential of underspecified structures* (pp. 179–192). Berlin: Walter de Gruyter.

Krajewski, G., A. L. Theakston, E. V. Lieven & M. Tomasello. (2011). How Polish children switch from one case to another when using novel nouns: Challenges for models of inflectional morphology. *Language and Cognitive Processes*, 26(4), 830-861.

Lapata, M., McDonald, S., & Keller, F. (1999). Determinants of Adjective-Noun Plausibility. Proceedings of EACL ’99, 30-36.

Levshina, N. (ms.). Associations between constructions and collexemes revisited: convergent evidence of divergent knowledge.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition,* 106, 1126–1177.

Lijffijt, J. & St. Th. Gries. (2012). Correction to "Dispersions and adjusted frequencies in corpora". *International Journal of Corpus Linguistics,* 17 (1), 147-149.

Mandelbrot, B. (1965). "Information Theory and Psycholinguistics". In B.B. Wolman and E. Nagel. *Scientific psychology*. Basic Books. Reprinted as Mandelbrot, B. (1968) [1965]. "Information Theory and Psycholinguistics". In R.C. Oldfield and J.C. Marchall. *Language*. Penguin Books.

McDonald, S.A., & Shillcock, R.C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. Language & Speech, 44(3), 295-322.

NKJP Polish National Corpus [http://www.nkjp.pl/]

Nagy, W.E., Anderson, R., Schommer, M., Scott, J.A. and Stallman, A. (1989) Morphological families in the internal lexicon. Reading Research Quarterly, 24(3), 263-282.

Newmeyer, F. (2003). Grammar is Grammar and Usage is Usage. *Language,* 79, 682- 707.

PELCRA Reference Corpus of Polish [http://korpus.ia.uni.lodz.pl/]

Polański, K. (1980-1990). *Słownik syntaktyczno-generatywny czasownków polskich I-IV.* Wrocław: Zakład Narodowy im. Ossolinskich

Polański, K. (1992). *Słownik syntaktyczno-generatywny czasownków polskich V.* Kraków: PAN.

Przepiórkowski, A. (2004). *Korpus IPI PAN. Wersja wstępna / The IPI PAN Corpus: Preliminary version*. Warszawa: IPI PAN.

Przepiórkowski, A. & A. Rosen. (2005). Czech and Polish Raising/Control with or without Structure Sharing. *Research in Language,* 3, 33-66.

Ramscar, M. and Yarlett, D. (2007). Linguistic Self-Correction in the Absence of Feedback: A New Approach to the Logical Problem of Language Acquisition. *Cognitive Science*, 31, 927–960.

Rao, R. P. N., B. A. Olshausen & M. S. Lewicki. (Eds). (2002). *Probabilistic Models of the Brain: Perception and Neural Function*. MIT Press.

Raymond, W. & E. Brown. (2012). Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. In Gries S. Th. and D. Divjak (Eds.) *Frequency effects in language representation* (pp. 35-52). Berlin: Mouton de Gruyter.

Recchia, G., Johns, B. T., & Jones, M. N. (2008). Context repetition benefits are dependent on context redundancy. In V. Sloutsky, K. McRae, & B. Love (Eds.), *Proceedings of the 30th Cognitive Science Society* (pp. 267-272).

Rescorla, R.A. & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black A.H. and W.F. Prokasy (Eds.) *Classical Conditioning II (*pp. 64–99). Appleton-Century-Crofts.

Rohde, D. (2003). *Improving the value of ratings by factoring out the generosity of raters*. [Available at http://tedlab.mit.edu/~dr/Papers/Rohde-Spindle.pdf. Last accessed 20 January 2012]

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science,* 12, 110-114.

Schmid, H.-J. (2000). *English abstract nouns as conceptual shells. From corpus to cognition*. Berlin - New York: Mouton de Gruyter. (Topics in English Linguistics 34).

Schmid, H.-J. (2010). *Does frequency in text instantiate entrenchment in the cognitive system?* In D. Glynn & K. Fisher (eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches* (pp. 101–133). Berlin: Mouton de Gruyter.

Schmid, H.-J. & H. Küchenhoff. (2013). Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics*, 24(3), 531-577.

Schmid, H.-J. & H. Küchenhoff. (2015). Reply to 'More (old and new) misunderstandings of collostructional analysis: on Schmid & Küchenhoff' by Stefan Th. Gries. *Cognitive Linguistics* 26 (3).

Schütze, C. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology.* Chicago: University of Chicago Press.

Schütze, C. & J. Sprouse (2013). Judgment data. In: R. J. Podesva & D. Sharma (eds.), *Research Methods in Linguistics* (pp. 27-50). Cambridge: Cambridge University Press.

Sedlmeier, P. & T. Betsch (eds.). (2002). *Frequency Processing and Cognition*. London: Oxford University Press.

Snider, N. & I. Arnon. (2012). A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon. In D. Divjak & S. Gries (Eds.) *Frequency effects in language representation* (pp. 127-164). Berlin: Mouton de Gruyter.

Solan, Z., D. Horn, E. Ruppin & Sh. Edelman. (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences,* 102, 11629–11634.

Sprouse, Jon. (2008). *The differential sensitivity of acceptability to processing effects. Linguistic Inquiry,* 39(4), 686-694.

Sprouse, J. (2013). *Acceptability Judgments*. *Oxford Bibliographies Online: Linguistics*. Edited by Mark Aronoff. [http://www.oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0097.xml]

Sprouse, J. & D. Almeida. (2013). The role of experimental syntax in an integrated cognitive science of language. In *The Cambridge Handbook of Biolinguistics* (pp. 181-202). Edited by K. Grohmann & C. Boeckx.

Staum Casasanto, L, Ph. Hofmeister & I. A. Sag. (2010). Understanding Acceptability Judgments: Additivity and working memory effects. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Sutton, R. S. & A.G. Barto. (1981). Towards a modern theory of adaptive networks: Expectation and prediction. *Psychological Review,* 88, 135-170.

Swan, O. (2002). A grammar of contemporary Polish. Bloomington, IN: Slavica.

Theakston, A. (2004). The role of entrenchment in children's and adults' performance on grammaticality judgment tasks. *Cognitive Development,* 19, 15–34.

Tomasello, M. (2007). Acquiring linguistic constructions. W. Damon & R.M. Lerner (eds), *Handbook of Child Psychology* (pp. 255-298). Hoboken, NJ: Wiley.

Townsend D. J., C. Carrithers & Th. G. Bever. (2001). Familial Handedness and Access to Words, Meaning, and Syntax during Sentence Comprehension. *Brain and Language,* 78 (3, September), 308-331.

Tremblay, A. (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. [Was available for download at http://www.ualberta.ca/~antoinet/ThesisDraft_10_B.pdf]

Ullman, M. T. & M. Walenski. (2005). Reply: Moving past the past tense. *Brain and Language,* 93, 248-252.

Yang, Ch. D. (2004). Universal Grammar, statistics, or both. *Trends in Cognitive Sciences,* 451-456.

Wiechmann, D. (2008). On the Computation of Collostruction Strength. *Corpus Linguistics and Linguistic Theory*, 4(2), 253-290.

Wiechmann, D. (2008). Sense-contingent lexical preferences and early parsing decisions. *Cognitive Linguistics*, 19 (3), 439-455.

Willems, R. M., S. L. Frank, A. D. Nijhof, P. Hagoort & A. van den Bosch. (2015). Prediction During Natural Language Comprehension. *Cerebral Cortex* Advance Access (published April 22, 2015; accessed 15 May 2015]

Zipf G. K. (1949). *Human behavior and the principle of least effort.* Addison-Wesley Press.

Appendix 1 – overview of the Polish verbs included in this study with English translation

Verb (ID)	**Verb** (imperfective_perfective aspect)	**Translation**	**Attested in *that*-clause**
1	zezwalać_zezwolić	Allow, permit, let	Yes
2	brzydzić się_	Abhor, loathe, have an aversion	No
3	przyrzekać_przyrzec	Promise	Yes
4	kochać_	Love	Yes
5	wzbraniać się_wzbronić się	Forbid	Yes
6	ośmielić się_ośmielić się	Venture, dare	No
7	zamyślać_zamyślić	Design	No
8	obawiać się_	Fear, be afraid, be anxious	Yes
9	umieć_	Know how, be able	No
10	starać się_postarać się	Endeavor, make efforts, take pain, try	Yes
11	decydować się_zdecydować się	Determine, decide	Yes
12	dawać się_dać się		No
13	pozwalać_pozwolić	Allow, permit, let	Yes
14	przyzwyczajać się_przyzwyczaić się	Become accustomed, get used	Yes
15	poczynać_począć	Begin, originate	No
16	zabraniać_zabronić	Forbid, prohibit, interdict	Yes
17	życzyć [sobie]_zażyczyć [sobie]	Wish, desire	Yes
18	kazać_kazać	Bid, order, let	Yes
19	proponować_zaproponować	Offer, propose	Yes
20	zakazywać_zakazać	Forbid, prohibit	Yes
21	móc_	Can, be able	No
22	poważać się_poważyć się		Yes
23	nawykać _nawyknąć	Become accustomed	Yes
24	pomagać_pomóc	Help, aid, assist	Yes
25	przysięgać_przysiąc	Swear	Yes
26	próbować_spróbować	Try, test, attempt	Yes
27	radzić_poradzić	Advise	Yes
28	dokańczać_dokończyć	Finish up, conclude	No
29	ślubować_ślubować	Vow, make a vow	Yes
30	uczyć się_nauczyć się	Learn	Yes
31	śpieszyć _pośpieszyć	Hurry, be in a hurry	Yes
32	ubóstwiać_	Idolize, adore	No
33	woleć_	Prefer	Yes
34	kończyć_skończyć	End, finish, conclude, close	No
35	_zechcieć	Become willing	Yes
36	godzić się_	Agree, consent	Yes
37	nienawidzić_	Hate, detest	No
38	pamiętać_	Remember, keep in mind	Yes
39	obiecywać [sobie]_obiecać [sobie]	Promise	Yes
40	_omieszkać	Fail	No

41	planować_zaplanować	Plan	Yes
42	mieć_	Have to	No
43	zobowiązywać się_zobowiązać się	Bind, pledge oneself	Yes
44	_uwziąć się	Set one's mind, become crazy	No
45	śmieć_	Dare, venture	No
46	dopomagać_dopomóc	Help, aid, assist	Yes
47	rozpoczynać_rozpocząć	Begin, start, commence	No
48	wstydzić się_	Be ashamed	Yes
49	zgadzać się_zgodzić się		Yes
50	kusić się_skusić się	Seek to obtain, attempt	Yes
51	zalecać_zalecić	Recommend, commend	Yes
52	zapominać_zapomnieć	Forget	No
53	krępować się_	Be embarrassed, feel uneasy	No
54	potrzebować_	Need, want, be in need of	Yes
55	bronić_	Defend, guard, vindicate, assert	Yes
56	raczyć_raczyć	Deign, condescend	No
57	silić się_	Make efforts, exert oneself	Yes
58	nakazać_nakazać	Order, command	Yes
59	zaczynać_zacząć	Begin, start, commence	No
60	bać się_	Be afraid of, fear	Yes
61	postanawiać_postanowić	Resolve, determine, make up one's mind	Yes
62	potrafić_potrafić	Know how to do, manage	No
63	uwielbiać_uwielbić	Adore, worship	Yes
64	musieć_	Be obliged to, have to	Yes
65	odważać się_odważyć się	Dare, venture	Yes
66	usiłować_	Make efforts, endeavor, attempt	Yes
67	ważyć się_odważyć się	Dare, venture	Yes
68	doradzać_doradzić	Advise	Yes
69	pragnąć_	Desire	Yes
70	zdążać_zdążyć	Manage to do (on time)	Yes
71	prosić_poprosić	Ask, beg, request	Yes
72	chcieć_	Want, be willing, intend, desire, wish	Yes
73	przyobiecywać_przyobiecać	Promise	Yes
74	polecać_polecić	Recommend	Yes
75	_zdołać	Be able	No
76	myśleć_	Think, mean	Yes
77	zamierzać_zamierzyć	Intend, mean, be going	Yes
78	wahać się_zawahać się	Hesitate, weaver	Yes
79	umożliwiać_umożliwić	Enable, make possible	Yes
80	lękać się_	Fear, be anxious	Yes
81	kwapić się_pokwapić się	Be eager	Yes
82	ofiarowywać się_ofiarować się		Yes

83	spodziewać się_	Hope, expect	Yes
84	uczyć_nauczyć	Teach, instruct	Yes
85	podejmować się_podjąć się	Undertake	Yes
86	kontynuować_	Continue	No
87	lubić_	Like, love	Yes
88	przestawać_przestać	Cease, stop, discontinue	No
89	szykować się_przyszykować się		Yes
90	przykazywać_przykazać	Order, command	Yes
91	_zaofiarować się		No
92	namawiać_namówić	Induce, persuade	Yes
93	rozkazywać_rozkazać	Order, command	Yes
94	przywykać_przywyknąć	Get accustomed to	Yes
95	żenować się_	Feel embarrassed	No

Appendix 2 – R code for the calculation of the association measures

```
a <- verb_in_construction_frequency
b <- verb_frequency - a
c <- construction_frequency - a
d <- total - a - b - c
n <- total_relevant_POS

# attraction or relative frequency of the verb in the construction
relfreq <- verb_in_construction_frequency/construction_frequency

# relative reliance
relrel <- verb_in_construction_frequency/verb_frequency

# deltaP with construction as cue and verb as response
dpc <- a/(a+c) -  b/(b+d)

# deltaP with verb as cue and construction as response
dpv <- a/(a+b) -  c/(c+d)

# logarithm of the odds ratio
logor <- log(a*d/b*c)

# discounted logarithm of the odds ration whereby 0.5 is added to each value to
avoid infinite values
logordisc <- log((a+0.5)*(d+0.5)/(b+0.5)*(c+0.5))

# logarithm of the relative risk, i.e. a ratio of the probability of the verb
occurring in the construction versus its chances of occurring in the other
constructions
logrelriskv <- log((a/(a+c))/(b/(b+d)))

# logarithm of the relative risk, i.e. a ratio of the probability of the
construction containing the verb versus its chances of containing the other
verbs
logrelriskc <- log((a/(a+b))/(c/(c+d)))

# pmi
pmi <- log(a/((a+b)*(a+c)/(a+b+c+d)))

# pmi squared
pmi2 <- log(a^2/((a+b)*(a+c)/(a+b+c+d)))

# pmi cubed
pmi3 <- log(a^3/((a+b)*(a+c)/(a+b+c+d)))

# local mutual information
lmi <- a*log(a/((a+b)*(a+c)/(a+b+c+d)))

# log-likelihood ratio
# first calculate the expected values aa, bb, cc, dd and then the log-
likelihood.
aa<- (a+b)*(a+c)/n
bb<-a+b-aa
cc<-a+c-aa
dd<-c+d-cc
loglik<-2*(a*log(a/aa)+b*log(b/bb)+c*log(c/cc)+d*log(d/dd))

# Dice coefficient
dice <- 2*a/((a+b)+(a+c))
```

```
# Jacard coefficient
jacard <- a/(a+b+c)

# minimum sensitivity
min <- min(a/(a+b), a/(a+c))

# geometric mean
geomean <- a/sqrt((a+b)*(a+c))

# logarithm of poisson
logPoisson <- log(exp(-(a+b)*(a+c)/n)*((a+b)*(a+c)/n)**a/factorial(a))

# poisson-stirling
PoissonStirling <- a*(log(a)-log((a+b)*(a+c)/n)-1)

# logarithm of the hypergeometric likelihood
loghyp<- log(phyper(a,a+c,b+d,a+b))
```