Usage frequencies of complement-taking verbs in Spanish

Paola E. Dussias[1,2], Alejandra Marful[3], Chip Gerfen[1,2], María Teresa Bajo Molina[3]
[1]Penn State University, [2]Center for Language Science, [3]Universidad de Granada

Abstract

Verb bias—or the tendency of a verb to appear with a certain type of complement—has been employed in psycholinguistic literature as a tool to test competing models of sentence processing. To date, the vast majority of sentence processing research involving verb bias has been conducted predominantly with monolingual English speakers. To test the generality of competing theories of sentence comprehension, it is important to conduct cross-linguistic studies of sentence processing and to add data from other languages to theories of sentence comprehension. Given this, it is critical for the field to develop verb bias estimates from speakers of languages other than English. Here, we report the results of a norming study for 135 Spanish verbs. One important goal of this study was to determine whether verb bias estimates remain stable over time, a question that to our knowledge has not been investigated. Our results demonstrate that individual verbs show significant similarities in their verb bias across the three years of data collection. To facilitate cross-linguistic work, we compare our verb bias results with those provided by monolingual English speakers in a previous norming study conducted by Garnsey, Lotocky, Pearlmutter, & Myers (1997).

1   Introduction

Verb bias norms have been critical for conducting studies in which the predictions of various models of sentence processing are tested. However, the interpretation of the results of such studies has been contentious. Findings that have been taken to reflect early influence of lexical information on syntactic decisions can also be explained in terms of reanalysis processes. This scenario is complicated by the recent evidence suggesting that different statistical analyses produce competing results (Kennison, 2009). To determine which sentence processing model best characterizes the cognitive architecture that underlies the ability to construct syntactic representations in real time, we need to combine existing theoretically sophisticated experimentation with on-line methods and statistical analyses that al-low us to unambiguously distinguish earlier- from later-stage processes of syntactic parsing. Until such methods are developed, converging evidence from cross-linguistic studies of monolingual sentence processing are critical for the construction of models of syntactic processing and for empirically testing the claims of each model. Up to now, such testing has come primarily from studies with monolingual speakers, and predominantly from studies with monolingual English speakers. Because verb bias provides a crucial testing ground for competing theories of sentence processing, it is critical for the field to develop verb bias norms in numerous languages. As our knowledge about the factors that modulate syntactic parsing expands, we need an increasingly rich set of norms in order to probe the emerging theoretical questions raised by the different models of sentence processing.

With this in mind, we report the results of a norming study in which the usage frequencies of 135 Spanish verbs were collected. Using verb bias data derived from normative studies to test competing models of sentence comprehension presupposes that norming results are consistent across time. To our knowledge, no study to date has explicitly addressed this question. Ensuring that verb specific biases are robust is important because past studies have shown the verb bias estimates are affected by a number of variables, including the method used to gather the data (e.g., sentence completion tasks or corpus-based approaches), the method used to compute verb bias (e.g., absolute frequency or relative frequency), and the specific senses of a verb (e.g., when 'conclude' is followed by a direct object, it usually means 'to bring to an end.' However, if followed by a subordinate clause, it can mean 'to arrive at an end by reasoning') (for an extensive discussion, see Gahl et al., 2004; Hare, McRae & Elman, 2004). In the present study, we examine the stability of verb bias estimates over time by collecting Spanish verb norms across three years.

2.   Method

2.1 Participants

A total of 575 monolingual speakers of Peninsular Spanish participated in the norming study. Participants were recruited over the course of

three years. They were students in the Department of Psychology at the University of Granada (Spain) and received course credit for their participation. All participants reported having minimal or no knowledge of a second language.

2.2 Materials

Eighty-one verbs were selected from a list of 100 English verbs used in a norming study conducted by Garnsey, Lotocky, Pearlmutter, & Myers (1997). The 81 verbs were translated into Spanish using the Collins Dictionary of Español-Inglés/English-Spanish (2000) and the resulting translations were subsequently verified by a Spanish-English translator. An additional 54 Spanish cognate verbs, selected from Nash's (1993) Spanish-English cognate dictionary, were also normed. Using these 135 verbs (81 + 54), two lists were created. List 1 contained the 81 target verbs and 49 fillers (e.g., dative verbs, intransitive verbs, and verbs that subcategorized for prepositional phrases) of similar length and number of syllables. Fillers were included in order to discourage participants from limiting their completions. List 2 included the 54 cognate verbs (useful to conduct research with bilinguals, given that cognates have a special status in the bilingual lexicon) and a subset of 46 verbs from the 81 verbs included in List 1. Because List 1 and List 2 would be administered to different groups of monolingual Spanish participants, the 46 verbs in List 2 were included to check for consistency in the participants' responses between the two lists.

Twenty randomized files were created, 10 for each list. Each file contained the target and filler verbs in their past tense form embedded in a sentence fragment headed by a subject (always a proper name).

2.3 Procedure

Usage frequencies were obtained using the sentence completion task described in Garnsey, Pearlmutter, Myers, & Lotocky (1997). Participants were instructed to read a sentence fragment silently and to fill in a completion by hand next to the corresponding verb. They were told that there were no constraints on the length of their completions and that the resulting sentence needed to be grammatically correct and semantically plausible. No other instructions were provided.

Data collection took place over the course of three years. Data for List 1 were collected twice during Year 1 and twice during Year 2. List 1 was administered to a total of 464 Spanish monolingual speakers. List 2 was administered to an additional group of 111 monolingual Spanish speakers during Year 3.

3. Results

Participants' responses to a verb were coded in three categories: Direct Object (DO) completions, Sentential Complement (SC) completions and Other completions. This last category included prepositional phrases, infinitivals, and completions headed by relative pronouns such as *lo que* (that which). For our analyses, we focused mainly on the DO/SC classification because of its theoretical importance in current sentence processing literature. Average responses in each category were computed. It was often the case that participants failed to provide a completion for a given verb, particularly if it occurred towards the bottom of the list. For these cases as well as for cases in which the responses were illegible, the trial was coded as missing, and the number of participants included for the particular verb was reduced by 1.

Because norms for the 81 Spanish verbs List 1 were collected at different times, Pearson r correlations were computed between DO average responses and between SC responses to determine whether the completions provided for each verb were consistent across time. Results showed a significant and positive correlation ($p < .0001$). In addition, we conducted a second correlation analysis that compared responses to the 46 verb entries that were common to List 1 and List 2. When responses were compared across the three different years in which the data were collected (i.e. Year 1 and Year 2 for List 1 and Year 3 for List 2), the results again showed a significant and positive correlation ($p < .0001$). Taken together, the findings suggest that participants' responses were highly consistent across time. Therefore, in subsequent analyses, we collapsed the responses for each verb.

Following a criteria frequently used in psycholinguistic verb bias studies (Trueswell et al., 1993; Garnsey, Pearlmutter, Myers, & Lotocky, 1997), a verb was classified as DO-bias if it was used at least twice as often with a direct object completion as with a sentential complement completion and as SC-bias if there were at least twice as many sentential complement completions as direct object ones. Verbs were classified as EQ-bias if the difference between DO and SC completions was not greater than 15%. Remaining verbs were classified as No Bias. We chose a

coding method that relied on relative frequencies (e.g., Garnsey, Lotocky, Pearlmutter, & Myers, 1997; Trueswell et al., 1993) rather than absolute frequencies (e.g., Merlo, 1994) for two reasons. First, we wanted to generate a set of verb bias norms in Spanish that would be comparable to existing English norms. Second, Gahl et al. (2004, Experiment 4) presented evidence suggesting that only the relative criterion for classifying verb bias could account for some of the results reported in a number of psycholinguistic studies. Therefore, we opted for coding our verbs using the relative method.

From the total 135 verbs normed, 50% were DO-bias, 23% were EQ-bias, 16% were SC-bias, and 11% were No-bias. Correlation analyses comparing the results of the 81 Spanish verbs to the equivalent English translations from the Garnsey et al. norms were positively significant (DO average, $r = .44$, $p < .0001$ and SC average, $r = .41$, $p < .001$). Establishing whether there are cross-linguistic differences in verb bias, particularly between SC-biased and DO-biased verbs in Spanish and English, is critical not only for conducting cross-linguistic studies, but also for identifying how the lexicon and grammar of a bilingual's two languages produce mutual influences and how competition between the two linguistic systems is resolved. Hence, a 2 (Language: Spanish vs. English) x 2 (Type: DO vs. SC) ANOVA on the average responses was computed. Neither a main effect of language ($F < 1$) nor an interaction between language and type ($F < 1$) were found. However, the results showed a main effect of type [$F(1,80) = 10.453$, $p < .01$] such that DO responses ($M = .46$, $SD = 32$) were more frequent than SC responses ($M = .31$, $SD = 23$), as shown by comparing responses in Spanish [$t(80) = 2.72$, $p < .01$], in English [$t(80) = 2.81$, $p < .01$], and when comparing Spanish to English [$t(80) = 2.88$, $p < .01$] and English to Spanish [$t(80) = -3.15$ $p < .01$].

When bias alone was taken into account, 49% of the verbs showed a different bias in Spanish and English. However, the vast majority of these involved cases in which a verb changed between EQ and either DO or SC bias. Reverses in bias between Spanish and English also occurred, but were considerably fewer.

4. Conclusion

A significant feature of our findings is the correlation in participants' responses across the different times data were collected. The implication of this finding for theory-building is encouraging because a highly reliable set of norms strengthens experimental findings relevant to theoretical issues being debated in the sentence processing literature. Another important result is the distribution of verb subcategorization frequencies between Spanish and English--only about half of the verbs normed share bias in the two languages. The availability of a corpus containing verbs with same and different bias will be of use to scholars who wish to conduct cross-linguistic studies of sentence processing. It will also enable researchers interested in bilingualism to examine lexical effects on sentence processing when bilinguals read in each of their two languages.

References

Gahl, S., Jurafsky, D., & Roland, D. (2004). Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, & Computers*, **36**, 432-443.

Garnsey, S. M., Lotocky, M. A., Pearlmutter, N. J., & Myers, E. M. (1997). *Argument structure frequency biases for 100 sentence-complement-taking verbs.* (Unpublished manuscript, University of Illinois at Urbana-Champaign).

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language*, **48**, 281-303.

Kennison, S. M. (2009). The use of verb information in parsing: Different statistical analyses lead to contradictory conclusions. *Journal of Psycholinguistic Research*, **38**, 363-378

Merlo, P. (1994). A corpus-based analysis of verb continua tion frequencies for syntactic processing. *Journal of Psycholinguistic Research*, **4**, 435-447.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence-processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 528-553.