

Processing Facilitation Strategies in OV and VO Languages: A Corpus Study

Luis Pastor, Itziar Laka

Department of Linguistics and Basque Studies, University of the Basque Country, Bilbao, Spain
Email: luis.pastor@ehu.es

Received June 25th, 2013; revised July 28th, 2013; accepted August 5th, 2013

Copyright © 2013 Luis Pastor, Itziar Laka. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The present corpus study aimed to examine whether Basque (OV) resorts more often than Spanish (VO) to certain grammatical operations, in order to minimize the number of arguments to be processed before the verb. Ueno & Polinsky (2009) argue that VO/OV languages use certain grammatical resources with different frequencies in order to facilitate real-time processing. They observe that both OV and VO languages in their sample (Japanese, Turkish and Spanish) have a similar frequency of use of subject *pro*-drop; however, they find that OV languages (Japanese, Turkish) use more intransitive sentences than VO languages (English, Spanish), and conclude this is an OV-specific strategy to facilitate processing. We conducted a comparative corpus study of Spanish (VO) and Basque (OV). Results show (a) that the frequency of use of subject *pro*-drop is higher in Basque than in Spanish; and (b) Basque does not use more intransitive sentences than Spanish; both languages have a similar frequency of intransitive sentences. Based on these findings, we conclude that the frequency of use of grammatical resources to facilitate the processing does not depend on a single typological trait (VO/OV) but it is modulated by the concurrence of other grammatical features.

Keywords: Basque; Spanish; Word Order; Pro-Drop; Language Processing; OV/VO Languages; Corpus Study

Introduction

A central question in contemporary linguistic theory is to what extent the properties of language result from conditions imposed by processing demands. Thus, for instance, The Minimalist Program (Chomsky, 1995) assumes as a premise that most properties of language derive from conditions imposed by the Conceptual-Intentional (C-I) and Articulatory-Perceptual (A-P) interfaces. Within psycholinguistics, the idea that processing requirements have a significant impact on the form of sentences was first explored in the *Depth Hypothesis* (Yngve, 1960) that provided an account for the different incidence of left-branching and right-branching structures in English based on processing constraints. A decade later, Bever (1970) argued that the form of language reflects general cognitive laws, in such a way that mechanisms of language processing are one of the factors determining grammatical phenomena. This issue has been investigated from different perspectives by many researchers, mostly in psycholinguistics, and it continues to be a much-researched area among linguists and psycholinguists; MacDonald (2013), for instance, currently argues that production demands heavily constrain linguistic output and typological choices.

The present paper seeks to contribute to our understanding of the interplay between grammar and the conditions imposed by the externalization of linguistic expressions, and it does so by exploring the hypothesis that facilitation of processing in real time guides the preferences with which languages with different basic word orders resort to certain grammatical phenomena.

Of the six possible basic word orders (SVO, SOV, VSO, VOS, OSV, OVS) (Greenberg, 1969), the vast majority of human languages are either VO (40.5%) or OV (35.4%) (Dryer, 2011). These basic word orders correspond strongly with other word order properties; in VO languages the genitive generally follows the noun, while in OV languages it almost always precedes it (Greenberg's *Universal 2*). Several studies have suggested that basic word order has an impact on language processing, i.e., that VO-OV languages tend to favor different syntactic constructions, because they are more efficient to process each language type (VO-OV). For instance, Hawkins (1994, 2004) argued that VO-OV grammars result from the optimization of processing requirements. Hawkins proposed a general principle, *Minimize Domains* (MiD), according to which the orderings with the smaller domain are preferred in processing. Thus, MiD predicts that VO languages will favor word orders where short constituents precede long ones, whereas OV languages will prefer to place long constituents before short ones, creating a mirror effect. Yamashita & Chang (2001) carried out an experimental exploration of Hawkins' hypothesis and found a preference in Japanese (OV) for placing long constituents before short ones. More recently, Ros, Laka, Fukumura, & Santesteban (2012) conducted a similar experimental exploration in Basque (OV) and also found a "long before short" preference. Both studies lend support to Hawkins' hypothesis that VO-OV languages employ different constituent ordering strategies in order to facilitate processing.

transitive verbs with a NP object, transitive verbs with a CP object and ditransitives (transitive verbs with direct and indirect NP objects) (Table 2).

Types of *pro*-drop were also recorded: for Basque, subject and object *pro*, and subject *pro* for Spanish (Table 3). When coding Basque sentences, both direct and indirect objects were included in the same “object” group.

In order to investigate the frequency of usage of *pro*-drop, a Pearson chi-square test was run to determine whether the distribution of sentences with and without *pro*-drop for either “Intransitive” or “Transitive” sentences for each language were significantly different. Details of *pro*-drop types (S[ubject]-drop and O[bject]-drop) were additionally recorded in order to examine whether either language showed a higher rate of *pro*-drop in transitive sentences as compared to intransitive sentences.

In order to investigate the frequency of usage of intransitive sentences (Intransitivity), “Intransitive” and “Transitive” occurrences were examined by means of a Pearson chi-square test in order to determine whether a significantly different distribution of these structures was found between the two languages. An

alpha level of .05 was used for all statistical tests.

Results

Results of the comparative corpus study reveal that in terms of the distribution of *pro*-drop, both languages make a significantly greater use of *pro*-drop in transitive than in intransitive sentences, thus replicating the findings in Ueno & Polinsky (2009). In Spanish, *pro*-drop is used in 42.5% of transitive sentences compared to 28.8% in intransitive sentences [χ^2 (I) = 28.626, $p < .001$]. In Basque *pro*-drop is used in 61.4% of transitive sentences against 19.9% of intransitive sentences [χ^2 (I) = 260.445, $p < .001$] (Figure 1). This finding is also true for each genre separately in both languages. In press, Spanish uses 34.7% of *pro*-drop cases in transitive sentences [intransitive vs. transitive: 15.3% vs. 34.7%, χ^2 (I) = 35.090, $p < .001$]; and Basque 49.4% [intransitive vs. transitive: 17.4% vs. 49.4%, χ^2 (I) = 83.122, $p < .001$]. In books, Spanish has 50.7% cases of *pro*-drop in transitives [intransitive vs. transitive: 41.6% vs. 50.7%, χ^2 (I) = 5.826, $p < .016$]; and Basque 71.6% [intransitive vs. transitive: 22.7% vs. 71.6%, χ^2 (I) = 178.871, $p < .001$].

Table 1.
Examples of Intransitive sentences in Spanish and Basque.

		Intransitive sentences							
		<i>Spanish</i>							
		S				V			
		[El	segundo	encuentro]	transita	por	los	mismos	derroteros
		the	second	meeting	go	along	the	same.PLUR	course.PLUR
Intransitive verbs		“The second meeting go along the same courses”							
		<i>Basque</i>							
		S				V			
		[Errepide	mapa	bat	lantzeko	unea]	[etorriko da]		
		road	map	one	prepare.to	moment. ABS	arrive.FUT AUX		
		“The moment to prepare a road map will arrive”							
		<i>Spanish</i>							
		S				Pred			
		Thor	no	[es	un	simple	superhéroe]		
		Thor	no	be	a	simple	superhero		
Non-verbal predicates		“Thor is not a simple superhero”							
		<i>Basque</i>							
		S				Pred			
		Eta	hau	ez	[da	planteamendu	estetiko	bat]	
		and this. ABS	no	be	approach	esthetic	one		
		“And this is not an esthetic approach”							
		<i>Spanish</i>							
		S				V			
		[Estas	rocas]	[son	transportadas]	en	dos	grandes	depósitos
		theses	rock.PLUR	be	transport.PART.PLUR	in	two	big.PLUR	tank.PLUR
				placed	in	the	shipdeck	place.PART.PLUR	in
				the	deck	of	the	ship	
Demotion structures		“These rocks are transported in two big tanks placed in the shipdeck”							
		<i>Basque</i>							
		V				S			
		Zelan	[diagnostikatzen	zen]	orain	arte	epilepsia		
		how	diagnose.PART	be	now	until	epilepsy.ABS		
		“How was diagnosed the epilepsy up to now”							

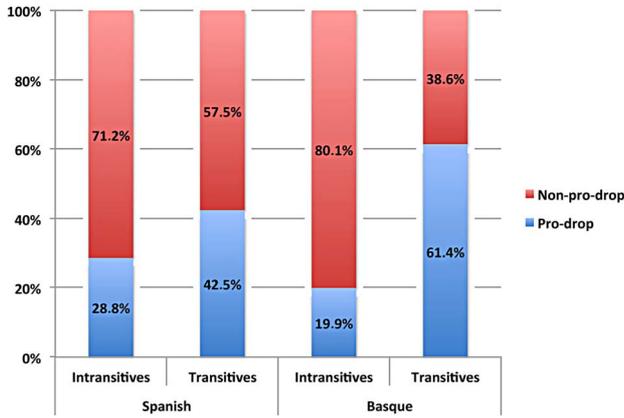


Figure 1. Distribution of *pro*-drop in intransitive and transitive sentences in Spanish and Basque.

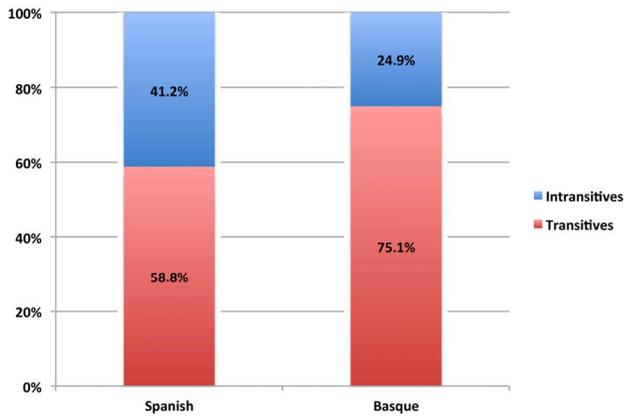


Figure 2. Distribution of subject *pro*-drop in intransitive and transitive sentences in Spanish and Basque.

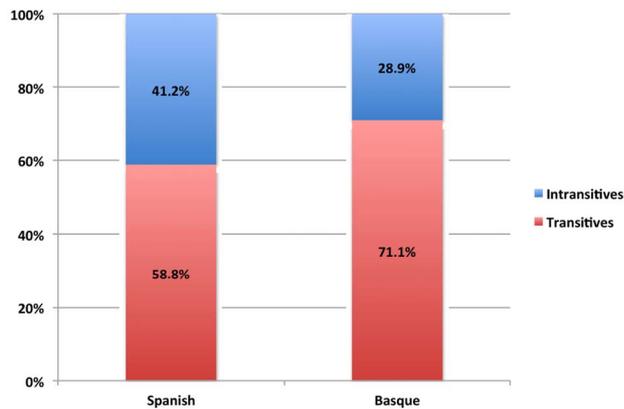


Figure 3. Distribution of subject *pro*-drop in intransitive and transitive sentences in Spanish and Basque.

tences, no significance differences obtain between Basque and Spanish. The Spanish corpus yields a 50.8% of intransitives sentences and the Basque corpus yields a 52.5% [Spanish vs. Basque: 50.8% vs. 52.5%, $\chi^2(1) = .824, p = .364$] (**Figure 4**). There are no significant differences in the analysis by gender: press [Spanish vs. Basque: 49.4% vs. 54.3%, $\chi^2(1) = 3.307, p$

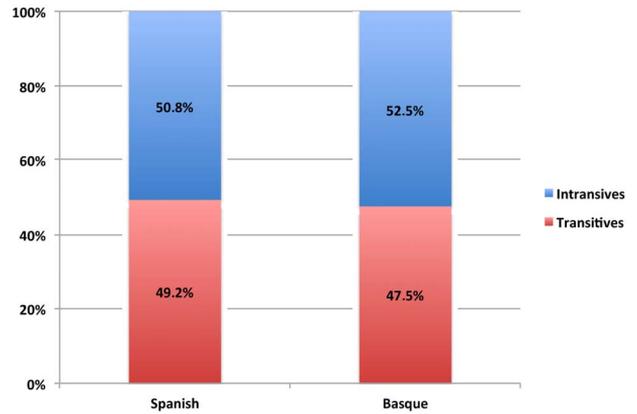


Figure 4. Distribution of intransitive and transitive sentences in Spanish and Basque.

$= .069$] and books [Spanish vs. Basque: 52.1% vs. 50.7%, $\chi^2(1) = .286, p = .593$].

Again, this pattern of results does not conform to the specific predictions in Ueno & Polinsky (2009) that OV languages should present a higher frequency of intransitive sentences than VO languages because intransitives minimize the number of arguments before the verb in these languages. However, the Basque data can be accounted for within the spirit of Ueno & Polinsky's general proposal, if it can be shown that Basque uses postverbal placement as an equivalent strategy to minimize the number of preverbal arguments and the processing cost entailed, a strategy not available to Japanese and Turkish that are strictly verb-final. In order to determine whether this is the case, we analyzed the frequency of postverbal arguments in declarative transitive sentences in our corpus. If transitive sentences with postverbal arguments outnumber those with both arguments before the verb, then it can be argued that the difference between Japanese and Turkish on the one hand, and Basque on the other in the relative frequency of intransitive sentences is related to the strict verb-final constraint in Japanese and Turkish, which is not operative in Basque.

Results show that in Basque only 27% of the transitive sentences have two arguments (XPXPV) in preverbal position against 73% with one argument (XPV) in the preverbal area [XPXPV vs. XPV: 27% vs. 73%, $\chi^2(1) = 82.368, p < .001$] (**Figure 5**).

As mentioned above, and unlike Japanese and Turkish, Basque grammar allows placing arguments in postverbal position (cf. (1c, d, e, f)), which we argue is a strategy to reduce the load in the preverbal area in transitive sentences. As shown in **Figure 6**, *pro*-drop (38%) and postverbal arguments (26.5%) are the strategies most frequently used in Basque in order to reduce preverbal arguments in transitive sentences [SOV vs. postverbal: $\chi^2(1) = 8.817, p < .003$].

Discussion

The present study compared Basque and Spanish written corpora in order to examine whether Basque, an OV language, resorts to *pro*-drop and intransitivity in order to reduce the number of arguments in the preverbal area, as predicted by Ueno & Polinsky for OV languages in general, or whether this

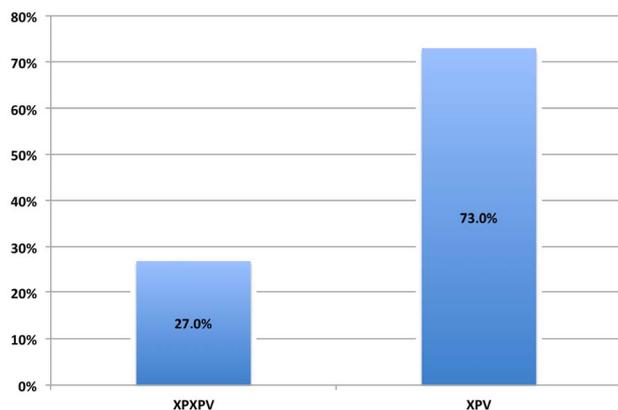


Figure 5. Distribution of transitive sentences with two (XPXPV) and one arguments (XPV) in Basque.

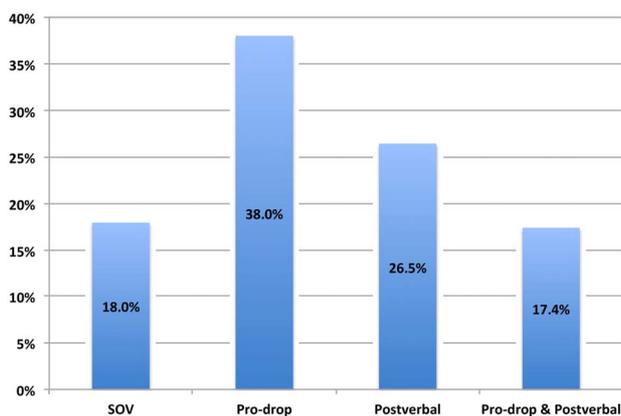


Figure 6. Distribution of *pro*-drop, postverbal and *pro*-drop/postverbal in transitive sentences in Basque.

language presents different strategies not available in previously studied OV languages.

Regarding *Pro*-drop, the results of the present corpus study show that the usage of *pro*-drop is more frequent in transitive sentences than in intransitive sentences for both Spanish and Basque, in accordance with the general findings in Ueno & Polinsky (2009). *Pro*-drop may facilitate processing because the elision of a pronoun requires less cognitive resources than expressing it phonologically (Yamashita, Chang, & Hirose, 2005). Moreover, null pronouns are easier to recover during the discourse since its references tend to have been mentioned previously (Kameyama, 1985, 1988; Walker, Iida & Cote, 1994; Turan, 1998; Prince, 1999). Within this general tendency, however, our study shows that Basque uses subject *pro*-drop more frequently than Spanish does. We argue this difference emerges from the difference in processing facilitation that results from subject-*pro* in each language. Recall that SOV sentences require to hold two arguments in memory before accessing the verb, while SVO sentences require to hold only one. Hence, omission of the subject would more effectively reduce the processing cost in SOV sentences than in SVO ones with respect to the point at which the verb is accessed. As regards Intransitivity, Ueno & Polinsky (2009) conclude that the more frequent usage of intransitive sentences is a specific strategy in OV languages to reduce processing cost. In OV languages a high usage of

intransitive sentences would minimize the short-term memory cost, since only one argument is retained in memory until processing the verb (Pritchett, 1992; Lindsley, 1975).

The results of the current study, however, reveal no significant difference neither in Basque (OV) or Spanish (VO) in the frequency of usage of intransitive sentences as compared to transitive sentences. When comparing Basque and Spanish collapsed, no significant difference was found in the usage intransitive sentences; both languages use intransitive sentences with similar frequency. The result obtained for Basque in the current study suggests that the high usage of intransitive sentence is not a specific strategy used by all OV languages to minimize the processing cost. Our results reveal that Basque uses postverbal placement as an equivalent strategy to reduce the number of preverbal arguments to be held in memory before the verb. Taking together the results of Ueno & Polinsky (2009) and the current study, we conclude that OV languages that do not allow arguments after the verb use more frequently intransitive sentences, whereas OV languages that allow arguments after the verb tend to use more postverbal arguments instead of intransitive sentences.

Conclusion

The present study shows that: a) both in Spanish and Basque *pro*-drop is used more frequently in transitive rather than intransitive sentences b) Basque resorts to subject *pro*-drop with a significantly higher frequency than Spanish does; c) there is no difference in frequency of usage of intransitive sentences in these two languages; and d) in Basque arguments tend to be placed in postverbal position as a main strategy to reduce the preverbal area.

The findings of this study provide support to the hypothesis that basic word order correlates with the frequency with which languages resort to certain syntactic phenomena. The frequency of usage of grammatical resources in order to facilitate the processing does not depend on a single typological feature (OV/VO) but seems to be modulated by the concurrence of other grammatical features, resulting in different profiles that depend on the parametric combination of each grammar.

Acknowledgements

This study has been supported by the grant BES-2010-030196 awarded to Pastor and the project (FFI2012-31360) funded by the Department of Science and Innovation (MICINN) of the Government of Spain, and the project IT665-13 funded by the Basque Government Funding is gratefully acknowledged.

REFERENCES

- Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279-362), New York, NY: Wiley and Sons.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- De Rijk, & Rudolf, P.G. (2007). *Standard Basque: A progressive grammar*. Cambridge, MA: MIT Press.
- Dryer, M. S. (2011). Order of subject, object, and verb. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library. <http://wals.info/chapter/81>
- Erdozia, K., Laka, I., Mestres-Misse, A., & Rodriguez-Fornells, A. (2009). Syntactic complexity and ambiguity resolution in a free word

- order language: Behavioral and electrophysiological evidences from Basque. *Brain and Language*, 109, 1-17. doi:10.1016/j.bandl.2008.12.003
- Garnsey, S. M., Pearlmutter, N. J., Myers, Elizabeth, & Lotockyc, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58-93. doi:10.1006/jmla.1997.2512
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of Language* (pp. 40-70). Cambridge, MA: MIT Press.
- Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, J. A. (2003). Efficiency and complexity in grammars: Three general principles. In J. Moore, & M. Polinsky (Eds.), *The Nature of Explanation in Linguistic Theory* (pp. 121-152). Stanford: CSLI Publications, Stanford University.
- Hualde, J. I., & Ortiz de Urbina, J. (2003). *A Grammar of Basque*. Berlin: Mouton de Gruyter. doi:10.1515/9783110895285
- Kameyama, M. (1985). *Zero anaphora: The case of Japanese*. Doctoral Dissertation, Stanford, CA: Stanford University.
- Kameyama, M. (1988). Japanese zero pronominal binding: Where syntax and discourse meet. In W. J. Poser (Ed.), *2nd International Workshop on Japanese Syntax* (pp. 47-73). Stanford, CA: CSLI Publications.
- Laka, I. (1996). *A brief grammar of Euskara, the Basque language*. www.ei.ehu.es/p289-content/eu/contenidos/informacion/grammar_euskara/en_doc/index.html
- Lindsay, J. R. (1975). Producing simple utterances: How far ahead do we plan? *Cognitive Psychology*, 7, 1-19. doi:10.1016/0010-0285(75)90002-X
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 1-16. doi:10.3389/fpsyg.2013.00226
- Pickering, M. J. (1993). Direct association and sentence processing: A reply to Gorrell and to Gibson and Hickok. *Language and Cognitive Processes*, 8, 163-196. doi:10.1080/01690969308406953
- Pickering, M. J., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, 6, 229-259. doi:10.1080/01690969108406944
- 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, and Cognition*, 19, 528-553. doi:10.1037/0278-7393.19.3.528
- Prince, E. F. (1999). Subject pro-drop in Yiddish. In P. Bosch, & R. van der Sandt (Eds.), *Focus: Linguistic, cognitive, and computational perspectives* (pp. 82-104). Cambridge, CA: Cambridge University Press.
- Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.
- Ros, I., Laka, I., Fukumura, K., & Santesteban, M. (2012). Long-before-short in head-final languages that agree. *The Architectures and Mechanisms for Language Processing Conference (AMLaP)*, Universidad di Trento.
- Turan, Ü. D. (1998). Ranking forward-looking centers in Turkish: Universal and language specific properties. In M. A. Walker, A. K. Joshi, & E. F. Prince (Eds.), *Centering theory in discourse* (pp. 139-161). Oxford: Clarendon Press.
- Ueno, M., & Polinsky, M. (2009). Does headedness affect processing? A new look at the VO-OV contrast. *Journal of Linguistics*, 45, 675-710. doi:10.1017/S0022226709990065
- Walker, M. A., Iida, M., & Cote, S. (1994). Japanese discourse and the process of centering. *Computational Linguistics*, 20, 193-231.
- Yamashita, H., & Chang, F. (2001). "Long before short" preference in the production of a head-final language. *Cognition*, 81, B45-B55. doi:10.1016/S0010-0277(01)00121-4
- Yamashita, H., Chang, F., & Hirose, Y. (2005). *Producers build structures only with overt arguments. The 18th Annual CUNY Conference on Human Sentence Processing*, Tucson.
- Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 140, 444-466.