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 25th, 2013; accepted August 2nd, 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.
Ueno & Polinsky (2009) tackle the interaction between basic word order (OV/VO) and processing constraints from a different but not incompatible perspective, and argue that VO-OV languages resort to different syntactic phenomena as strategies to facilitate processing, and that these distinct patterns are reflected in the frequency with which they employ these grammatical resources.

Ueno & Polinsky (2009) follow the assumption that the verb plays a pivotal role in sentence processing, because argument relations are resolved when the verb is reached (Head-driven model, Pritchett, 1992; Pickering & Barry, 1991; Pickering, 1993; Trueswell, Tanenhaus, & Kello, 1993; Garney, Pearlmuter, Myers, & Lotocky, 1997). According to this assumption, SOV sentences should present an extra processing cost because two arguments must be retained in working memory before processing the verb: the subject (S) and object (O). Ueno & Polinsky (2009) hypothesize that OV languages will resort to some syntactic phenomena (such as pro-drop, intransitive and others) more frequently than VO languages in order to minimize the amount of arguments expressed before the verb. In particular, they argue that pro-drop (unexpressed arguments) and Intransitivity might serve as strategies to reduce the number of preverbal arguments in OV languages. The former minimizes the number of arguments phonologically expressed, whereas the latter minimizes the number of arguments required by the verb.

To test this hypothesis, Ueno & Polinsky (2009) conducted two comparative corpus studies. A first study compared English (VO) and Japanese (OV) written sentences, and it revealed a similar frequency in the usage of pro-drop in transitive sentences in both languages, and a more frequent use of intransitive sentences in Japanese (OV) as compared to English (VO). The second study compared the narrative production of native speakers of English, Spanish (VO), Japanese and Turkish (OV), and it replicates the results obtained in the first study: pro-drop is used more often in transitive than in intransitive sentences in all languages (Spanish, Japanese and Turkish) and intransitive sentences occur more frequently in OV languages (Japanese and Turkish) than in VO languages (English and Spanish). Thus, whereas pro-drop is more frequently used in transitive sentences in both VO-OV languages with equivalent frequency, intransitive sentences are significantly more frequent in OV languages. Therefore, according to this study, resort to intransitive forms reflects an OV-specific strategy, directed at minimizing the number of preverbal arguments in a sentence.

**The Current Study**

The present investigation extends Ueno & Polinsky’s (2009) study, by comparing Spanish (VO) and Basque (OV). We conducted a comparative corpus study of Spanish (VO) and Basque (OV) aimed at determining whether Basque, presents a higher frequency of intransitives as do Turkish and Japanese in comparison to VO languages like Spanish, while the frequency of pro-drop would be similar in Basque and Spanish.

One main reason to pursue this study is to ascertain the general validity of the results obtained by Ueno & Polinsky for OV languages, and to explore Spanish in a larger corpus. Both Spanish and Basque can place subjects and objects after the verb, a possibility not available in English, Turkish, Japanese and Korean. Furthermore, in the present study, the corpora employed are significantly larger than those used by Ueno & Polinsky (2009).

**Characteristics of Basque**

Like Japanese and Turkish, Basque has basic SOV word order (cf. 1a). However, unlike Japanese and Turkish, which have strictly verb-final sentences, Basque may have postverbal arguments (cf. (1c, d, e, f)) (Laka, 1996; Hualde & Ortiz de Urbina, 2003; De Rijk, 2007; Erdozia et al., 2009).

1) a) gizon-a-k emakume-a-Ø ikusi du (SOV) man-Det-ERG woman-Det-ABS see AUX “The man has seen the woman”
   b) emakumea gizonak ikusi du (OSV) c) gizonak ikusi du emakumea (SVO)
   d) emakumea ikusi du gizonak (OV)
   e) ikusi du gizonak emakumea (VSO)
   f) ikusi du emakumea gizonak (VOS)

In addition, Basque is a pro-drop language, i.e., the arguments that agree with the verb do not have to be phonologically expressed. Both subjects and objects can be omitted (cf. (2h, c, d)), unlike in Spanish, where there only subjects can be unexpressed.

2) a) Ni-k zu-ri liburua-Ø ekarri dizut I-ERG you-DAT book-ABS bring AUX “I bring you the book”
   b) pro zuri liburua ekarri dizut [subject pro-drop]
   c) Nik pro liburua ekarri dizut [direct object pro-drop]
   d) Nik zuri pro ekarri dizut [indirect object pro-drop]
   e) pro pro pro ekarri dizut [subject, indirect object and direct object pro-drop]

These two grammatical properties of Basque make it a relevant grammar to investigate whether all OV resort to similar strategies to facilitate processing. In particular, we seek to ascertain whether Basque behaves like Japanese and Turkish using a high frequency of intransitive sentences in order to reduce arguments of the preverbal area, or whether this language will resort to postverbal arguments as a specific strategy, not available to Japanese and Turkish. We also seek to determine the status of pro-drop as a facilitating strategy, by comparing the ratios of pro-drop in Basque and Spanish.

**Method**

A total of 2800 sentences from press and books in both Spanish and Basque (1400 sentences/source × 2 sources) were analyzed for sentence type. Each source was divided into different genres in both languages in order to have a heterogeneous corpus. Press source was divided in 7 genres (business, culture, society, politics, world, national and sports) (100 sentences/genre × 7 genres × 2 languages); whereas books in 4 genres (comedy, mystery, historical and non-fictional) (175 sentences/genre × 4 genres × 2 languages).

Matrix clauses of the sentences were manually coded for sentence type, while adjoined and embedded clauses were disregarded. They were classified as either “Intransitive” or “Transitive”. The Intransitive category included intransitive verbs, non-verbal sentences (adjectival and nominal), and demotion structures (passives with or without the by-phrase, impersonals, middle voices...) (Table 1). The transitive category included...
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 “In-transitive” 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 ($\chi^2 (1) = 28.626, p < .001$). In Basque pro-drop is used in 61.4% of transitive sentences against 19.9% of intransitive sentences ($\chi^2 (1) = 260.445, p < .001$) (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%, $\chi^2 (1) = 35.090, p < .001$]; and Basque 49.4% [intransitive vs. transitive: 17.4% vs. 49.4%, $\chi^2 (1) = 83.122, p < .001$]. In books, Spanish has 50.7% cases of pro-drop in transitives [intransitive vs. transitive: 41.6% vs. 50.7%, $\chi^2 (1) = 5.826, p < .016$]; and Basque 71.6% [intransitive vs. transitive: 22.7% vs. 71.6%, $\chi^2 (1) = 178.871, p < .001$].

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

<table>
<thead>
<tr>
<th>Intransitive sentences</th>
<th>Spanish</th>
<th>Basque</th>
</tr>
</thead>
<tbody>
<tr>
<td>El segundo encuentro transita por los mismos derroteros</td>
<td>“The second meeting go along the same courses”</td>
<td></td>
</tr>
<tr>
<td>Thor no es un simple superhéroe</td>
<td>“Thor is not a simple superhero”</td>
<td></td>
</tr>
<tr>
<td>Estas rocas son transportadas en dos grandes depósitos situados en la cubierta de la nave</td>
<td>“These rocks are transported in two big tanks placed in the shipdeck”</td>
<td></td>
</tr>
<tr>
<td>Zelan diagnostikatzen zen orain arte epilepsia</td>
<td>“How was diagnosed the epilepsy up to now”</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.
Examples of Transitive sentences in Spanish and Basque.

<table>
<thead>
<tr>
<th>Transitive sentences</th>
<th>Spanish</th>
<th>Basque</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitives</td>
<td>[El edificio] tiene [una antigüedad aproximada de 70 años]</td>
<td>[Herri honek] [premia izugarria] du, people this.ERG necessity much.ABS have</td>
</tr>
<tr>
<td></td>
<td>“The building has an antiquity of approximately 70 years”</td>
<td>“This people have a lot of necessity”</td>
</tr>
<tr>
<td></td>
<td>[El presidente Mohamed Hosni Mubarak] [ha decidido] [renunciar a su cargo de presidente de la República]</td>
<td>[Nik beti [esaten dut] [hau opari gisa hartzen dugula,] L.ERG always say AUX this gift as AUX.take</td>
</tr>
<tr>
<td></td>
<td>“The president Mohamed Hosni Mubarak has decided to resign his post of president of the Republic”</td>
<td>“I always say that we take this as a present”</td>
</tr>
<tr>
<td>Ditransitives</td>
<td>[Esa formula] [ha permitido] [a inversores profesionales y a empresas de gran tamaño] allow.PART to investor.PLUR professional.PLUR and to company.PLUR big size</td>
<td>[Alemaniarrek [20 milioi euro] [zor dizkiote] [lantegi honi] German.PLUR.ERG 20 million euro.ABS owe AUX factory this.DAT</td>
</tr>
<tr>
<td></td>
<td>“That formula has allowed professional investors and companies of big size to achieve a certain degree of range in their operations”</td>
<td>“Germans owe this factory 20 million euros”</td>
</tr>
</tbody>
</table>

When the usage of pro-drop in transitive sentences is compared in both languages, significantly more instances of pro-drop are found in Basque (75.1%) than in Spanish (58.8%) [Spanish vs. Basque: 58.8% vs. 75.1%, \( \chi^2 \) (I) = 32.791, \( p < .001 \)] (Figure 2).

This result departs from the general finding in Ueno & Polinsky (2009). This difference in the distribution of pro could be due to the fact that in Basque three argument-types can be unexpressed (subject, object and indirect object), whereas in all the languages studied by Ueno & Polinsky (2009) were like Spanish, in that they only have subject pro-drop. In other words, Basque would present a higher frequency of unexpressed arguments because it can omit them in more cases. This explanation predicts that the frequency of subject pro should be equivalent in Basque and Spanish. However, when only subject pro-drop is considered, Basque continues to show a significantly higher frequency of pro-drop in transitive sentences (71.2%) than Spanish (58.8%) [Spanish vs. Basque: 58.8% vs. 71.2%, \( \chi^2 \) (I) = 16.477, \( p < .001 \)] (Figure 3).

Regarding the distribution of intransitive vs. transitive sen-
tences, no significance differences obtain between Basque and Spanish. The Spanish corpus yields a 50.8% of intransitive sentences and the Basque corpus yields a 52.5% [Spanish vs. Basque: 50.8% vs. 52.5%, $\chi^2$ (I) = .824, $p = .364$] (Figure 4).

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.

I 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.

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.

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%, $\chi^2$ (I) = 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: $\chi^2$ (I) = 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
null pronouns are easier to recover during the dis-
pressing it phonologically (Yamashita, Chang, & Hirose, 2005).

elision of a pronoun requires less cognitive resources than ex-
-drop may facilitate processing because the
linsky (2009).

Basque, in accordance with the general findings in Ueno & Po-
show that the usage of
matical features, resulting in different profiles that depend on
languages resort to certain syntactic phenomena. The frequency
languages present different strategies not available in previous-
Regarding Pro-drop, the results of the present corpus study
show that the usage of pro-drop is more frequent in transitive
than in intransitive sentences for both Spanish and
Basque, in accordance with the general findings in Ueno & Po-
linsky (2009). Pro-drop may facilitate processing because the
elision of a pronoun requires less cognitive resources than ex-
pressing it phonologically (Yamashita, Chang, & Hirose, 2005).

Moreover, null pronouns are easier to recover during the disc-
ourse since its references tend to have been mentioned pre-
viously (Kameyama, 1985, 1988; Walker, Iida & Cote, 1994;
Turan, 1998; Prince, 1999). Within this general tendency, how-
ever, our study shows that Basque uses subject pro-drop more
frequently than Spanish does. We argue this difference emerges
d from the difference in processing facilitation that results from
subject-pro in each language. Recall that SOV sentences re-
quire 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 pro-
cessing 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 pro-
cessing the verb (Pritchett, 1992; Lindsey, 1975).

The results of the current study, however, reveal no signifi-
cant 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 col-
lapsed, no significant difference was found in the usage tran-
sitive 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 intransi-
tive sentences.

**Conclusion**

The present study shows that: a) both in Spanish and Basque
pro-drop is used more frequently in transitive rather than in-
transitive sentences b) Basque resorts to subject pro-drop with a
significatively 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 proc-
cessing does not depend on a single typological feature (OV/VO)
but seems to be modulated by the concurrence of other gram-
matical 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**

Hayes (Ed.), Cognition and the development of language (pp. 279-
362), New York, NY: Wiley and Sons.

Press.

De Rijk, & Rudolf, P.G. (2007). Standard Basque: A progressive gram-

Dryer, M. S. (2011). Order of subject, object, and verb. In M. Haspel-
math, 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


