**Aiming at shorter dependencies: The role of agreement morphology**

Idoia Ros[1], Mikel Santesteban[1], Kumiko Fukumura[2] and Itziar Laka[1]

[1]Department of Linguistics and Basque Studies, University of the Basque Country (UPV/EHU)

[2]School of Psychological Sciences and Health, University of Strathclyde (UK)

Number of words:

Main text: 7,657

Main text + References + Acknowledgements: 10,542

Corresponding author: Idoia Ros

Department of Linguistics and Basque Studies

Psycholinguistics Laboratory

University of the Basque Country UPV/EHU

Unibertsitateko Ibilbidea 5

Vitoria-Gasteiz, 01006

Spain

Tel. +34 945 014 452

Fax: +34 945 013 200

Email: idoia.ros@gmail.com

**Abstract**

This study examined word order preferences as a function of phrasal length in Basque. Basque is an OV-language with flexible sentence word order and rich verb agreement. Contrary to the universal short-before-long preference predicted by availability models, Hawkins has argued that short-before-long orders are preferred in VO-languages such as English, whereas long-before-short orders are preferred in OV-languages such as Japanese (Hawkins, 2004). However, it is unclear how length affects word order preferences when an OV-language has rich verb agreement and allows postverbal arguments. We found a general long-before-short preference, and a tendency to place the verb in a sentence-medial position when one constituent is long. We argue that since agreement morphology signals the thematic role and case of surrounding phrases, it contributes to speeding up sentence processing. We conclude that morphologically rich languages employ both general adjacency mechanisms and language-specific resources to enhance language efficiency.

**Keywords:**

*Introduction*

In the language sciences, the impact of processing demands on sentence word order is a central topic of research. In the last decades, linguistics has come to accept that processing constraints play a role in word order preferences (Chomsky, 1995; Dryer, 1980, 1992; Hawkins, 1983, 1994; Wasow, 2002), thus converging with psycholinguistic models that have tackled the question of how processing demands impact on sentence word order variations (Arnold, Wasow, Asudeh, & Alrega, 2004; Bock, 1982; Bock & Levelt, 1994; Bresnan, Cueni, Nikitina, & Baayen, 2007; Ferreira & Dell, 2000; Gibson, 1998; among others), including accounts based on information theoretic considerations (Gibson, Piantadosi, Brink, Bergen, Lim, & Saxe, 2013; Maurits, Perfors, & Navarro, 2010, inter alia). Thus, there is general agreement that sentence word order is modulated by factors external to the grammar, although there are differing views on the specifics of this interaction. This study explores how phrasal length affects sentence word order in Basque, a heavily inflected OV language.

The question of exactly how processing constraints underlie word order preferences in sentence production has remained controversial to date. *Availability-based models* of sentence production claim that word order variations are primarily affected by ease of production, whereas *efficiency-based theories* focus more on ease of comprehension. In availability models, speakers favour sentence structures where more readily available words are placed earlier, to reduce the need of holding already-retrieved information in memory and to start producing early (Bock & Irwin, 1980; Bock & Warren, 1985; Branigan, Pickering, & Tanaka, 2008; Bresnan et al., 2007; Ferreira & Dell, 2000; Ferreira & Yoshita, 2003; Levelt & Maassen, 1981; Prat-Sala & Branigan, 2000; Tanaka, Branigan, McLean, & Pickering, 2011; see Jaeger & Norcliffe, 2009, for an overview). In contrast, efficiency-based theories argue that the parser prefers word orders that allow the fastest computation of constituent structure (Hawkins, 1994, 2014; Temperley, 2007).

Specifically, Hawkins put forth a model that explains how speakers preferentially choose word orders that can optimally facilitate parsing, which should in turn influence production preferences.

The current study tests these contrasting hypotheses by investigating the impact of length (operationalised as word count[1]) on the production of transitive and ditransitive sentences in Basque, for which availability-based and efficiency-based models make different predictions. As we discuss later in more detail, *availability-based accounts* predict a general preference to place phrases that are *short* before those that are *long*, independently of the basic word order or other typological properties of the language (Arnold, Wasow, Losongco, & Ginstrom, 2000; de Smedt, 1994; *Easy First principle* in MacDonald, 2013; Stallings & MacDonald, 2011; Stallings, MacDonald, & O'Seaghda, 1998; Wasow, 1997a, 1997b).  In contrast, in Hawkins' *Performance Theory of Order and Constituency model* (henceforth*, PTOC)*, word order preferences depend on phrasal length and head direction of the language in question (whether dependents follow or precede heads in phrases). In Hawkins' theory, head-initial (VO) languages will prefer short-before-long orders, while head-final (OV) languages will favour long-before-short orders.

Based on our results for Basque, we argue that VO and OV languages have different word order preferences for sentences involving long constituents. Specifically, speakers of both rigid and flexible OV languages (for a classification of OV languages in terms of word order rigidity, see Aldai, 2011) favour long-before-short orders, as predicted by Hawkins but contrary to availability-based models that predict a short-before-long preference. We also argue that, on top of head direction (head initial-VO/head final-OV), other grammatical properties, such as verb agreement, modulate sentence word order preferences. In particular, we argue that agreement can act as an argument recognition device to speed up sentence processing.

In the following sections, we will first discuss how investigating Basque can enhance our understanding of the effect of length on word order preferences, and we then describe the relevant linguistic characteristics of Basque. Secondly, we contrast the availability-based

accounts and the *PTOC* model, and their respective predictions for Basque word order

preferences, before reporting the results of our experiment.

*Why Basque?*

Psycholinguistic studies have explored word order variations as a function of length in VO

and OV languages (Stallings et al., 1998; Yamashita & Chang, 2001, among others).

Investigating the effect of length in OV languages is particularly important because, as mentioned

earlier, availability-based accounts and Hawkins' efficiency-based *PTOC* model make different

predictions. Unlike availability models, Hawkins' *PTOC* model predicts that OV languages will

show a preference for placing long constituents before short ones. There is written corpus data

supporting this prediction for Turkish, Japanese and Korean (Choi, 2007; Hakuta, 1981; Hawkins,

1994) and psycholinguistic evidence from two OV languages: Japanese (Yamashita & Chang,

2001) and Korean (Dennison, 2008). Basque is also an OV language, but it differs from Japanese

and Korean in two (probably related) properties. First, whereas Japanese and Korean are strictly

verb-final (Hinds, 1986; Sohn, 1994), Basque allows postverbal constituents. Second, both

Japanese and Korean lack verb agreement (Hinds, 1986; Sohn, 1994), whereas Basque has

obligatory verb agreement with subject, object and indirect object NPs. Thus, Basque allows us to

explore the impact of length in word order preferences in a non-rigid OV language that has verb

agreement.

*Basque*

Basque is an OV language isolate with rich inflectional morphology, which allows both

verb-medial and verb-final word order. It is spoken in the Basque Country (north-eastern Spain

and south-western France) by about 700,000 people.

Basque presents a number of typological properties that are not found in other European languages: ergative alignment; rich agglutinative morphology; overt case on NPs; subject, object and dative verb agreement. Additionally, subjects, objects and datives can all be freely omitted (it is a three-way *pro-drop* language) (de Rijk, 2007; Hualde & Ortiz de Urbina, 2011). Given the agglutinative nature of its morphology, agreement morphemes are fully decomposable (Laka, 1993): verbal inflection agrees with the object (O), the subject (S) and the indirect object (IO) if there is one (see examples (1)-(4)). These agreement morphemes indicate whether these phrases are first, second or third person, singular or plural. Although basic word order is S(IO)OV (de Rijk, 2007; Elordieta, 2001), Basque grammar allows significant freedom in the order of phrases in a sentence. In particular, constituents can be placed either before or after the verb, as illustrated in sentences (1)-(4)[2], where all variants are truth-conditionally equivalent. Example (1) represents canonical word order, commonly used when the sentence denotes discourse-new information. Other phrase orders are also possible, as shown in (2)-(4).

(1) **S-IO-O-V**

*Andre-a-k          gizon-a-ri        egunkari-a            ekarri      d-i-o-ø*

Woman-Det$_{sg}$-S   man-Det$_{sg}$-IO  newspaper-Det$_{sg}$.O  broughtAux$_{root-3sg.O-3sg.IO-3sg.S}$

The woman has brought the man the newspaper

(2) **S-O-IO-V**

*Andre-a-k         egunkari-a            gizon-e-i        ekarri        d-i-e-ø*

Woman-Det$_{sg}$-S  newspaper-Det$_{sg}$.O  man-Det$_{pl}$-IO   brought       Aux$_{root-3sg.O-3pl.IO-3sg.S}$

(3) **IO-O-V-S**

*Gizon-a-ri       egunkari-ak              ekarri        d-izki-o-ø                        andre-a-k*

Man-Det$_{sg}$-IO    newspaper-Det$_{pl}$.O    brought    Aux$_{root-3pl.O-3sg.IO-3sg.S}$    woman-Det$_{sg}$-S

(4) **O-IO-V-S**

*Egunkari-a          gizon-a-ri      ekarri    d-i-o-te          andre-e-k*

Newspaper-Det$_{sg}$.O    man-Det$_{sg}$-IO    brought    Aux$_{root-3sg.O-3sg.IO-3pl.S}$    woman-Det$_{pl}$-S

*Availability-based models*

  Although most research on availability-based production has focused on the effects of conceptual accessibility (Bock, 1982; Bock & Warren, 1985; Ferreira, 1996; Ferreira & Dell, 2000, among others), several researchers have proposed a general preference to place short phrases before long ones (Arnold et al., 2000; de Smedt, 1994; *Easy First principle* in MacDonald, 2013; Stallings et al., 1998; Stallings & MacDonald, 2011; Wasow, 1997a, 1997b). The idea is that sentences such as (5b) are preferred over sentences such as (5a) below, because shorter phrases (*to Bill*) should be generally easier for speakers to retrieve from memory than longer phrases (*the book she had been searching for since last Christmas*) and hence can be assembled earlier. This preference to produce shorter phrases before longer ones should hold regardless of the typological properties of the languages involved. Indeed, the tendency to shift long constituents over short ones (often known as *heavy-NP shift*) is widely attested in corpora studies in VO languages like English, German[3], Hungarian, Greek, Polish, Finnish, Rumanian and Russian (de Marneffe, Grimm, Arnon, Kirby, & Bresnan, 2012; Hawkins, 1994; Hoffman, 1999; Kizach, 2012; Köhler, 1999; Lohse, Hawkins, & Wasow, 2004; Temperley, 2007; Uszkoreit, 1987; Uszkoreit et al., 1998; Wasow, 2002; Wiechmann & Lohmann, 2013; for an overview, see Jaeger & Norcliffe, 2009). There is also evidence from language production experiments supporting a short-before-long preference in English (Arnold et al., 2000; Marblestone, 2007; Stallings et al., 1998; Stallings & MacDonald, 2011).

However, not all languages have demonstrated short-before-long preferences. As mentioned earlier, in Japanese and Korean, speakers tend to favour long-before-short word orders. Thus, some authors have suggested that the effect of availability is modulated by language-specific characteristics (Wasow, 2002, 2013; Yamashita & Chang, 2001; Chang, 2009). Yamashita and Chang, who found a long-before-short preference in Japanese, argued that word order preferences are affected by both conceptual saliency and ease of formulation. Although long phrases are more complex to formulate than shorter phrases, they are conceptually more salient because the additional words make the semantic representation of the head noun richer and stronger. English has a very strict word order, such that form-related factors exert a stronger effect, and because short phrases are easier to formulate than longer phrases, they tend to be placed before longer ones. In contrast, Japanese is a scrambling language with null pronouns (Saito, 2004), and that allows speakers to produce salient arguments early. Therefore, according to Yamashita & Chang' proposal, the freer the word order, the stronger the effects of conceptual saliency, and hence the tendency to place longer constituents before shorter ones (but see Tanaka, Branigan, & Pickering, 2011).

Wasow (2002, 2013) also suggested that the general preference to place short elements before long ones could compete with other constraints. That is, in VO languages the short-before-long word order might arise not only from the preference to place shorter elements earlier in the sentence, but also from the preference to avoid complex constituents in the middle of the sentence, which are more costly to process (Chomsky, 1965; Gibson, 2000; Miller & Chomsky, 1963). For instance, (5a) is harder to process than (5b) below, because the relative clause embedded within the direct object increases the distance between the verb and the indirect object, making the integration of the indirect object harder. Therefore, speakers may prepose the indirect object to reduce the processing cost. In strictly OV languages, however, these two constraints pull in different directions: a short-before-long linearisation of constituents places short

elements earlier in the sentence, but it also creates a center-embedded structure. In these cases, speakers should seek to avoid center-embedding by shifting longer constituents to a sentence-initial position, leading to a long-before-short ordering. This hypothesis about the opposing word order pattern in VO and OV languages is fairly similar to Hawkins' *Performance Theory of Order and Constituency*, which we now discuss in detail.

*The Performance Theory of Order and Constituency (PTOC) model*

The *PTOC* model (Hawkins, 1994, 2009, 2014) comprises three general principles that predict word order preferences in different languages: *Minimise Domains*, *Maximise On-line Processing* and *Minimise Forms*. The first two principles are relevant to the current study.

The first principle that determines word order preferences is the *Minimise Domains principle* (henceforth, MiD). Both phrasal length and the head-direction of the language (OV vs. VO) determine word order preferences. Word orders that require fewer computations to reach a parsing decision about all the syntactic and semantic relations between the constituents within minimal processing domains are ranked as more efficient and hence more frequently produced[4]. The calculation of the size of a processing domain provides us with a metric that makes precise predictions regarding ranked frequencies of competing structures. This metric consists of the sum of the number of immediate constituents (ICs) divided by the sum of the number of non-immediate constituents (or words). This ratio quantifies the number of words that need to be processed in order to identify the immediate constituents of a given phrase, and thus represents the processing efficiency of a given linguistic sequence. The higher the ratio, the more efficient a processing domain. Hawkins predicts that the word orders with the optimal ratios will be produced more frequently than non-optimal orders. For instance, in the example sentences in English (5) and Japanese (6) below, (5a) and (6a) represent *non-shifted* canonical word orders, in which the direct object precedes the indirect object (5a) or follows the indirect object (6a). In contrast, (5b)

and (6b) represent *shifted* word orders, where the direct object is moved after (5b) or before (6b) the indirect object. MiD predicts that in both (5) and (6), the shifted orders (b) will be preferred over the non-shifted orders (a) (examples taken from Hawkins, 2004: 26 and Yamashita & Chang, 2001: B52).

(5a) I [VP **gave** [NP **the book she had been searching for since last Christmas**] [PP **to** Bill]]

        1    2   3   4   5   6   7    8   9  10   11   12

ICs: 3, {V, PP, NP}

Number of words: 12, {gave, the, book, she, had, been, searching, for, since, last,

  Christmas, to}

*MiD ratio: 3/12 = 25%*

(5b) I [VP **gave** [PP **to Bill**] [NP **the** book she had been searching for since last Christmas]]

        1    2  3    4

ICs: 3, {V, PP, NP}

Number of words: 4, {gave, to, Bill, the}

*MiD ratio: 3/4 = 75%*

(6a) Masako-wa[VP [NP **otoko-ni**][NP **sinbun-de syookai-sarete-ita okasi-o**] **todoketa**]

          1         2         3         4    5

Masako-Top [VP [NP man-Dat] [NP newspaper-in appeared cake-Acc] delivered.]

Masako delivered the cake [which was] introduced in the newspaper to the man

ICs: 3, {V, NP, NP}

Number of words: 5, {otokoni, sinbunde, syookaisareteita, okasio, todoketa}

*MiD ratio: 3/5 =60%*

(6b) Masako-wa[$_{VP}$ [$_{NP}$ sinbun-de syookai-sareteita **okasi-o**][$_{NP}$ **otoko-ni] todoketa**]

<div style="text-align:right">
|———————|———————|
1        2        3
</div>

Masako-Top [$_{VP}$ [$_{NP}$ newspaper-in appeared cake-Acc] [$_{NP}$ man-Dat] delivered.]

Masako delivered the cake [which was] introduced in the newspaper to the man

ICs: 3, {V, PP, NP}

Number of words: 3, {okasio, otokoni, todoketa}

*MiD ratio: 3/3 = 100%*

Consistent with Hawkins' predictions, corpus and experimental studies have consistently

found a short-before-long preference in VO languages (e.g., English), but a long-before-short

preference in OV languages (e.g., Japanese, Korean and Turkish) (Choi, 2007; Kondo &

Yamashita, 2011; Hakuta, 1981; Hawkins, 1994). Two production experiments provide evidence

for a long-before-short tendency in Korean and Japanese (Dennison, 2008; Yamashita & Chang,

2001). Further support for the *PTOC* hypothesis comes from computational models that showed

that shorter dependencies increase parsing accuracy (Collins, 2003; Eisner & Smith, 2005;

Sleator & Temperley, 1991; Temperley, 2008).

    As discussed above, among the various hypotheses put forth, only the *PTOC* model

straightforwardly predicts a long-before-short preference for Basque. Following the way in which

Hawkins calculates efficiency metrics (1994), the verb-participle and the inflected auxiliary count

as two separate words in our calculations. Recall that, according to the *PTOC*, the higher the ratio

of a given sentence in comparison to others, the more efficient its processing will be and hence

the more frequently it will be produced. The efficiency ratios for some possible sentences in

Basque are shown in Tables 1 and 2. According to these ratios, in Basque, in a sentence with a

long object phrase, shifted orders where the object appears before shorter constituents are

optimally efficient. Therefore, we should expect O-S-V and O-V-S in transitive sentences and O-

IO-S-V in ditransitive ones to be produced most frequently.

The second principle that guides word order preferences in the *PTOC* model, the

*Maximise On-line Processing principle (*henceforth, MaOP)*, is concerned with speed in

communication. It seeks to provide the earliest possible access to linguistic representation by

avoiding processing delays and ambiguities. The general assumption is that many syntactic and

semantic properties are not assignable to positions lower in thematic and syntactic hierarchy

positions independently from the higher positions (agent > recipient  > patient; c-commanding >

c-commanded) (Primus, 1999) and that these assignment constraints will influence processing

preferences, such that word orders that allow these assignments with minimum delays are

preferrable. MaOP predicts a general preference for subjects to precede objects and agents to

precede patients, because a patient requires a co-occurring agent, on which it is thematically

dependent, and a c-commanded element requires a c-commanding one, on which it is

syntactically dependent. Also, MaOP predicts that when the dependent category (patient, O)

precedes the independent one (agent, S), both positions will be as adjacent as possible, because

any intervening elements (e.g., V) will cause a delay in thematic assignment. Note that

processing dependencies between O and V are assumed to be symmetric: O is dependent on V

because it receives its thematic role from V, and V is dependent on O for semantic and syntactic

disambiguation (i. e*., John ran vs. ran the race* vs. *ran the water* vs. *ran the advertisement*, from

Keenan, 1979, as cited in Hawkins, 2014). Therefore, no preference for either S-O-V or S-V-O is

predicted based on MaOP.

According to MaOP, word orders with non-shifted NPs (e.g., S-O; IO-O) are favoured,

because shifted NPs alter the preferred word orders for property assignment. Therefore, in

Basque either S-V-O or S-O-V in the transitive sentences and either S-IO-O-V or S-IO-V-O in the

ditransitive sentences are considered optimal. Importantly, MaOP also predicts that if NPs are

shifted, verb-final word orders, where the verb does not intervene between the noun phrases, should be preferred over verb-medial ones. This is in order to avoid any extra delay in the assignment of properties caused by the intervening V. Therefore, in Basque S-O-IO-V will be preferred over other possible orders such as S-O-V-IO or O-V-S-IO in the ditransitive sentences and O-S-V over O-V-S in the transitive ones (see Table 3).

In short, for Basque, MiD predicts a preference for long constituents to be shifted before shorter ones, whereas MaOP predicts a preference for verb-final orders over verb-medial ones when NPs are shifted. One important and unexplored question is how the two principles, MiD and MaOP, interact with each other (Hawkins, 2009). One possibility is to assume that if different word orders have approximately the same MiD ratio, then the word order with a better MaOP ratio should be preferred (Hawkins, 2004). The resulting ranking for the type of sentences used in our experiment is shown in Table 3, where predictions derived solely from MiD or MaOP are also included. Briefly, even if both principles interact, long-before-short orders are still favoured over short-before-long ones (O-S over S-O, and O-IO over IO-O, when the object is long), but shifted long-before-short verb-final orders will be ranked higher than shifted long-before-short verb-medial ones (O-S-V over O-V-S; O-IO-S-V and S-O-IO-V over O-V-S-IO).

---------------------------------------------- Table 1 around here -----------------------------------------------
---------------------------------------------- Table 2 around here -----------------------------------------------
---------------------------------------------- Table 3 around here -----------------------------------------------

Given that the crucial element to detect long-before-short preferences in our experiment is the object phrase, only calculations for sequences containing a long object have been included in the tables above.

*On the role of agreement morphology in efficient processing*

Hawkins (2014) notes that, apart from adjacency, there may be other ways of enhancing efficiency. The MaOP principle predicts that syntactic processing in VO languages will tend to rely on agreement because it provides early access to the argument structure before all NPs are accessed. On the other hand, in OV languages, the parser will tend to rely on case-marking because it provides early online information about syntactic case and thematic role before the verb. However, NPs in Basque can be omitted or may be ambiguous because of case syncretism (see Erdocia, Laka, Mestres-Missé, & Rodríguez-Fornells, 2009). Thus, case marking alone does not always allow one to straightforwardly determine sentence structure. However, the early encounter of the verbal complex (lexical verb + auxiliary) can compensate for this. Specifically, Pablos (2011) has argued for Basque that agreement information provided by the auxiliary[5] acts as a cue that assists the parser in predicting upcoming elements. Thus for instance, in a sentence like (7) the inflected auxiliary (highlighted) provides information about the upcoming indirect object *gizon-a-ri* "to the man" and object *egunkari-a* "the newspaper".

(7)

| *Andre-a-k* | *ez* | ***d-i-o-ø*** | *gizon-a-ri* | *egunkari-a* | *ekarri* |
|---|---|---|---|---|---|
| Woman-Det$_{sg}$-S | Neg | Aux$_{root-3sg.O-3sg.IO-3sg.S}$ | man-Det$_{sg}$-IO | newspaper-Det$_{sg}$.O | brought |

The woman did not bring the man the newspaper

Placing the inflected verb early in the sentence facilitates the identification of grammatical functions and semantic roles of surrounding constituents, because agreement obligatorily signals the number and type of arguments the sentence contains. We argue that when speakers produce constituent structures that are less preferrable according to the MaOP principle (e.g., O-S over S-O, and O-IO over IO-O), placing the lexical verb and the agreement carrying auxiliary between

the two constituents facilitates rather than delays the assignment of syntactic and thematic

properties and thus this verb-medial order is preferred over immediate adjacency of shifted

constituents. This is because less linguistic material needs to be accessed to determine *initial*

sentence structure than in the case of shifted verb-final structures, all other things being equal.

For instance, in a long-before-short verb-medial ditransitive sentence in Basque (long-NP

verb+auxiliary short-NP short-NP), when reaching the verbal complex, the previous absolutive-

marked long constituent can be unambiguously identified as an object. Furthermore, the

presence of the subject and the indirect object will be clearly indicated by verb-agreement, even

before encountering the subject and indirect object phrases. In a long-before-short verb-final

ditransitive sentence, sentence structure cannot be resolved until all constituents of the sentence

have been accessed.

Note that non-rigid SOV languages can access initial structure either by case-marking

(while shifted constituents remain adjacent to each other, consistent to MaOP predictions

discussed earlier) or by placing the verb in a medial position (so that verb-argument structure is

assigned immediately, by virtue of the agreement markers). Hawkins' proposal does not make

any clear-cut prediction as to whether the early position of the verb should be favoured as more

efficient than case marking in languages of this type. Thus, in addition to whether Basque deploys

the same long-before-short preference as do rigidly verb-final SOV languages, we also examine

whether verb-medial structures are preferred over verb-final ones when noun phrases have been

reordered due to their length.

### *The experiment*

We carried out a cued recall production-task. Participants were presented with phrases

on a computer screen and were asked to construct sentences using these phrases. We

manipulated the length of the phrases (short vs. long) by adding relative clauses to different

constituents. We then investigated the effect of constituent length on sentence word order. Specifically, we looked at the likelihood that speakers would produce shifted (O-S; O-IO) and non-shifted canonical word orders (S-O; IO-O) in transitive and ditransitive sentences.

Hawkins' *PTOC* model predicts that speakers of any type of OV languages, in order to improve efficiency, will favour long-before-short word orders. Thus, we expect Basque speakers to show a larger preference to shift the object before other constituents when it is long (long-O) than when it is short (short-O and a short- or long-S or a short- or long-IO), which would result in long-O preceding S in transitive sentences and IO in ditransitive ones. In contrast, if, as predicted by availability-based models, short phrases are easier to retrieve from memory than longer ones, we then expect Basque speakers to show a stronger preference to shift the object after other constituents when it is long (long-O) than when it is short (short-O, and a short- or long-S, and a short- or long-IO), which would result in long-O following S in transitive sentences and following IO in ditransitive ones.

Finally, if agreement morphology plays a role in efficient processing, as suggested above, we would expect speakers to produce more verb-medial sentences in contexts where they also produced more shifted orders than in contexts where they produced non-shifted orders. Thus, if the predictions of the *PTOC* apply, this will occur in sentences containing a long-O, and if the predictions of availability-based models apply, this will occur in sentences containing a long-S for transitives or a long-IO for ditransitives).

*Method*

*Participants.* Twenty-five native speakers of Basque (15 females, mean age: 19.39) participated in the experiment. They were all undergraduate students at the University of the Basque Country (UPV/EHU). Adult speakers of Basque are bilingual with Spanish or French. All participants in this study are native speakers of Basque, and have Spanish as their second language (see Table 4).

Data from one participant was excluded from the analysis due to technical problems with the

recorder.

--------------------------------------------- Table 4 around here ----------------------------------------------------

*Materials.* Two sets of experimental sentences were created, following the design in Yamashita

and Chang (2001). For the first set, 24 transitive sentences were generated. Three different

versions of each sentence were created. We attached a relative clause to (a) no constituent, or

(b) the subject or (c) the direct object. This yielded the three main conditions of the Length

variable: (a) short subject and short direct object condition (All-Short); (b) long subject and short

direct object condition (Long-S); (c) short subject and long direct object condition (Long-O) (see

Table 5).

For the second set, 30 ditransitive sentences were created. As before, we created three

versions of each sentence, attaching a relative clause to (a) no constituent, (b) the indirect object

or (c) the direct object. This yielded the three main conditions of the Length variable: (a) short

direct object and short indirect object condition (All-Short); (b) short direct object and long indirect

object condition (Long-IO); (c) long direct object condition and short indirect object condition

(Long-O) (see Table 6).

To avoid effects due to reading habits, we manipulated the location on the screen of the

noun phrases of the target sentences. For the transitive sentences, in one case the subject was

placed in a box in the upper-right corner of the screen and the object was in the box in the

bottom-left corner (S-O), whereas in the other case the positions were reversed (O-S). Similarly,

for the ditransitive sentences, in one condition the indirect object was placed in a box in the

bottom-left corner of the screen and the direct object was in the box in the bottom-right corner

(IO-O), whereas in the other condition the positions were reversed (O-IO). In both transitive and

ditransitive sentences the verb remained in the upper-left corner.


------------------------------------------ Table 5 around here --------------------------------------------------

------------------------------------------ Table 6 around here --------------------------------------------------


*Procedure.* We constructed six lists containing 54 experimental items (24 transitive and 30

ditransitive experimental items), and 56 fillers. Across the six lists, all six versions of a given item

set were presented once. Each participant was presented with one list, in an individually

randomized order, with the constraint that no two experimental sentences of the same set were

presented next to each other. The experiment was programmed with DMDX (Forster & Forster,

2003).

Participants were tested individually. They first pressed the spacebar, which triggered the

display of the first screen with four boxes containing sentence components. Participants prepared

a sentence using the components given and pressed the spacebar when they were ready. After a

blank screen for 500 milliseconds a simple arithmetic problem followed. The delay and the

arithmetic problem were created in order to encourage participants to produce the sentences

from their meaning, rather than by covert rehearsal (Yamashita & Chang, 2001). Participants

pressed the spacebar when they solved the arithmetic problem. At this point, the verb of the

sentence to be recalled was presented in the middle of the screen as a cue, and participants

produced the sentence. Unlike in Yamashita and Chang (2001), the verb was presented in the

middle of the screen, and it was not surrounded by empty boxes so as not bias any strategic

responses.

Having produced the sentence orally, participants were prompted to respond to whether

the sentence had appeared before in the experiment by pressing either Yes (key Z) or No (key

M). Twenty of the 56 fillers were presented twice for this recognition task. The position of the verb in the fillers varied with respect to experimental items and location of verb and arguments in the fillers was randomized. Experimental sessions were digitally recorded and transcribed for analysis.

*Scoring.* Participants' responses were scored in terms of (i) word order (*shifted* versus *non-shifted*) and (ii) verb position (*medial* versus *final*). We scored *non-shifted order* when sentences were appropriately recalled and the constituents were arranged in their canonical word order (Transitive sentences: S-O; Ditransitive sentences: IO-O). We scored *shifted order* when sentences were appropriately recalled and the constituents were arranged in a non-canonical word order (Transitive sentences: O-S; Ditransitive sentences: O-IO). The position of the verb was coded as *medial* when the verb appeared between other constituents and as *final* when it appeared at the end of the sentence.

We scored responses as *miscellaneous* when sentences were not recalled at all (n = 14, 1.1%) or recalled only partially, resulting in ungrammatical utterances (n = 12, .9%), or when participants swapped the original length of the constituents (n = 1, .1%). These miscellaneous responses were excluded from further analysis. In order to avoid data loss, however, responses in which participants substituted a word with a related one were included in the main analyses (n = 125, 9.6%) (e.g., *proposamena* - "proposal" - instead of *dokumentua* - "document"). Our scoring resulted in 1,269 (97.9%) target responses.

*Results.* We analysed the number of *shifted response*s (relative to *non-shifted responses*) for transitive and ditransitive responses separately and the number of *verb-medial responses* (relative to *verb-final responses*) for ditransitive sentences only (the model was not run for transitive sentences for lack of sufficient data). Because our dependent variables were binary, we analysed all our data using logit mixed effects models (Breslow & Clayton, 1993, as cited in Jaeger, 2008). In all models we included Length as a fixed effect. We collapsed across Position of

the constituents in the Screen because it was a counterbalancing variable. Importantly, when the

inclusion of the Position in the Screen predictor (i.e., in the main model of ditransitive sentences[6])

yielded a better fit with the data, the results' significance of our critical Long-O condition were not

altered in any of the models (all $p$s < .02). We always used the maximum random effect structure

justified by the data by using model comparisons (Baayen, Davidson, & Bates, 2008). All data

were analysed using lme4 package (Bates, Maechler & Bolker, 2011) of the statistical software R

(version 2.14.2: R Development Core Team, 2012). First, we will report our analysis for transitive

responses.

*Transitive responses.* Table 7 reports the frequency and mean percentages of shifted and non-

shifted responses and miscellaneous responses, as well as verb-medial and verb-final

responses. In this model our dependent variable was the production of shifted vs. non-shifted

word orders. The Length predictor was Helmert-coded to avoid collinearity effects (after Helmert

contrasts, there was no indication of collinearity, $r$s < .02). By-participant and by-item random

intercepts were included. Table 9 reports the results. The Long-S contrast of the model in Table 9

compared the mean of All-Short with the mean of Long-S, and the Long-O contrast compared the

mean of Long-O with the mean of All-Short and of Long-S.


-------------------------------------------- Table 7 around here ----------------------------------------------
-------------------------------------------- Table 8 around here ----------------------------------------------


        The results of the main analysis revealed a significant negative intercept indicating that

participants produced more non-shifted (S-O) orders than shifted orders (O-S) when both

constituents were short. Also, the non-significant coefficient for the Long-S Length predictor

revealed that participants produced a similar number of non-shifted responses (S-O) in both the

Long-S and the All-Short conditions. Importantly, the significant positive coefficient for the Long-O

Length predictor indicated that participants were more likely to produce the long object earlier in the sentence compared to a short constituent of the same type (All-Short, Long-S), which resulted in more shifted orders (O-S). However, we acknowledge that the high log-odds in the coefficients indicate that the model might be overfitted due to the relatively limited number of observations per cell (Jaeger, 2011), so our results should be interpreted with caution.

*Ditransitive responses.* The means are reported in Table 8. Note that all sentence word orders with less than 5% of occurrences in all conditions are grouped under *Others*. Although there were some miscellaneous responses (16, 2.2%), they were equally distributed across Length conditions  ($p > .4$). Length was Helmert-coded to reduce collinearity (after Helmert contrasts, high collinearity levels ($r$s > . 9) were reduced to mild-collinearity levels ($r$s. > .6)). By-participant and by-item random intercepts and a by-participant random slope for Length were included as random effects. Table 9 reports the results. The Long-IO contrast of the main model in Table 9 compared the mean of All-Short with the mean of Long-IO, and the second contrast compared the mean of Long-O with the mean of All-Short and of Long-IO.

The results of the main analyses revealed a non-significant intercept indicating that participants produced shifted orders (O-IO) and non-shifted orders (IO-O) similarly often when both constituents were short. The negative coefficient for the Long-IO Length predictor revealed that participants produced fewer shifted orders (O-IO) in the Long-IO condition than in the All-Short condition. That is, they were more likely to produce the IO phrase earlier in the sentence when the IO phrase was long (long-IO) than when it was not (All-Short)[6]. Importantly, the significant positive coefficient for the Long-O Length predictor indicated that participants were more likely to produce the long object (Long-O) earlier in the sentence compared to a short constituent of the same type (All-Short, Long-IO), which yielded more shifted orders (O-IO). We also ran a model that directly compared the preference of speakers to produce shifted orders in Long-O vs. All-short condition ($p < . 01$).

We then modelled the production of *verb-medial* and *verb-final* responses with the aim of exploring whether the production of shifted structures also led to the production of verb-medial over verb-final sentences. Given that verb-medial responses were generally rare, we could not meaningfully analyse if verb position was conditional on word order shift. We thus indirectly examined this issue by analysing the log-odds of verb-medial and verb-final responses as a function of Length in the ditransitive model, where we had enough responses. Because there was no sign of collinearity in the ditransitive model ($r$s < .01), Length was effect-coded. By-participant and by-item random intercepts and a by-participant random slope for Length were included as random effects (see Table 9).

The results of the verb-position analysis revealed a non-significant intercept indicating that participants produced verb-medial and verb-final responses similarly often. The non-significant coefficient for the Long-IO Length predictor revealed that participants did not produce more verb-medial sentences when the IO phrase was long (Long-IO) than when it was not (All-Short). In contrast, the significant positive coefficient for the Long-O Length predictor indicated that participants produced more verb-medial responses in sentences containing a long object (Long-O) than when they contained a short subject and a short object (All-Short condition).


---------------------------------------------- Table 9 around here -----------------------------------------------


In sum, our results reveal a tendency to use non-shifted canonical orders, although this was significant only for transitive sentences. Also, when Basque speakers do resort to shifted word orders, they prefer to place long constituents before short ones, both in transitive and ditransitive sentences. Additionally, evidence from the production of ditransitive sentences showed that, in those contexts in which speakers showed a preference to produce shifted word

orders (i.e., in Long-O conditions), they also showed a preference to place the verb earlier in the

sentence (i.e., in a verb-medial position).

**Discussion**

In this study we examined the effect of phrasal length on sentence word order

preferences in a flexible sentence word order OV language (Basque), a language that has both

case-marking and rich verb agreement and allows for both verb-medial and verb-final orders. We

found that Basque speakers fronted the object more frequently when it was long than when it was

short, revealing a general preference to produce long phrases before short ones, as has been

found in rigidly verb-final OV languages (Dennison, 2008; Yamashita & Chang, 2001). We can

thus conclude that speakers of OV languages do not comply with the general short-before-long

preference predicted by availability-based models (Arnold et al., 2000; *Easy First principle* in

MacDonald, 2013; Stallings et al., 1998; Stallings & MacDonald, 2011; Wasow, 1997a, 1997b).

Our findings support the prediction made by the *PTOC* model that speakers of OV languages will

favour long-before-short sentence word orders (Hawkins, 2004, 2014). That is, the short-before-

long preference is exclusive to VO languages.

Some of our findings were not predicted by the *PTOC*. In the remainder of this paper, we

first discuss these findings. Lastly, we consider the implications of our data viv à vis some other

accounts, namely MacWhinney's *Competition Model* (1987) and Yamashita and Chang's (2011)

proposal, which argues that the effects of conceptual accessibility might vary cross-linguistically.

*Why verb-medial responses?*

Regarding the specific ranking of sentence word order preferences, the *PTOC*

predictions were not met. According to *PTOC*, Basque speakers should prefer verb-final

sentences over verb-medial ones in ditransitive sentences when the object is long and has been

shifted. However, our results reveal a preference for verb-medial sentences in this case. One

possible explanation is that speakers produce the verb early to compensate for processing

difficulties for the listener. As discussed earlier, verb agreement directly signals thematic

assignment in Basque and can guide the exact prediction of upcoming sentence structure, so that

when the verb is fronted, agreement morphology provides enough information to fully determine

sentence structure (see Pablos, 2011 for a similar claim). Therefore, the position of the verb is

likely to affect processing efficiency ratios in Basque, thus impacting word order preferences.

An alternative explanation for the preference for verb-medial sentences in Basque might

be MacDonald's *similarity-based interference* (MacDonald, 2013). It has been shown that when

semantically related nouns are planned and uttered close to each other, production difficulty

increases (Smith & Wheeldon, 2004), possibly because speakers need to inhibit one of the nouns

when producing the other noun but later have to retrieve it. MacDonald (2013) argues that this

similarity-based interference may influence the choice of utterance form. Speakers may choose

passive subject relatives more than active object relatives when two animate nouns are to be

mentioned than when one noun is animate, and the other inanimate (Gennari et al., 2012). We

may therefore wonder whether Basque speakers tend to produce verb-medial responses more

when they have to produce similar animate NPs, so that they can avoid producing the similar

nouns adjacently. However, although an interference-based account may explain why speakers

sometimes produced transitive OVS sentences ($NP_{animate}$-V-$NP_{animate}$) in Basque, it is unclear how

it can explain verb-medial responses for ditransitive sentences, where two animate NPs appear at

the same side of the verb (OVSIO: $NP_{inanimate}$-V-$NP_{animate}$-$NP_{animate}$). Verb-medial structures were

more frequent in our study for ditransitive sentences than for transitive ones, so it seems unlikely

that the preference for the verb position would result from similarity-based interference. Most

importantly, this account does not straightforwardly explain why verb-medial responses were

affected by the length of the constituents (i.e., in the case of transitive sentences no verb-medial

responses were produced when all the NPs were short in All-short conditions regardless of NP animacy).

Furthermore, the *Noisy-Channel account* (Gibson et al., 2013) may explain why Basque speakers sometimes produced SVO utterances. According to this account, SVO arises as a result of communicative pressures, due to users' sensitivity to the possibility of noise corrupting the linguistic signal. When a potential ambiguity arises in an SOV word order (e.g., in the case of semantically reversible events, where two possible agents precede the verb) a shift to SVO is predicted in order to maximise the listener's ability to recover the meaning, given the noisy process. Therefore, this account predicts that, in semantically reversible sentences like the transitive ones in our study, where case marking is ambiguous (an absolutive singular NP can be an intransitive S or an O), the number of SOV word order sentences will decrease and SVO orders will increase. However, our participants rarely produced SVO orders. They produced SOV orders 87.9% of the time and only produced shifted word orders 12.1% of the time. This general preference for canonical word orders in transitive sentences in Basque is consistent with corpus studies (de Rijk, 1969/1998). It is also possible that the hypothesized shift from SOV to SVO is restricted to situations where case marking is completely ambiguous and inefficient to disambiguate the interpretation of reversible events. Alternatively, related communicative efficiency accounts that do not attribute confusability primarily to argument order confusion (e.g., Fedzechkina, Jaeger, & Newport, 2012) predict SVO order in cases when the properties of the arguments would make OSV a plausible alternative order (Kurumada & Jaeger, 2013). This is compatible with our results.

This overall preference for SOV order even when sentences contain long phrases, unexpected under Hawkins' account, might result from a grammaticalization process (see Hawkins, 2014 for related discussion). Future research will have to explore whether the pattern of responses we obtained, clearly favouring SOV over SVO responses, might have been

conditioned by the experimental paradigm and whether more naturalistic set-ups would yield

more SVO sentences (Lockridge & Brennan, 2002).

*The Competition Model: A trade-off between processing facilitation cues*

The effect of length on word order appears to be smaller in Basque than in other

languages in similar experimental conditions: the rate of NP-shift in transitive sentences was only

16.1% in Basque, whereas it reached 26% and 28.2% in Japanese and Korean respectively

(Dennison, 2008; Yamashita & Chang, 2001). In ditransitive sentences (for which no Korean data

exist), Japanese also resorted to NP-shift more often (40.3%) than Basque (17.9%). In a study on

English involving long and short PPs, Stallings and MacDonald (2011) reported length affected

word order in about 40% of cases. These differences may be due to the less reliable  role that

sentence word order plays as a processing cue in Basque.

MacWhinney's *Competition Model* (1987) argues that the relative strength of processing

preferences depends on the availability of cues that signal grammatical structure (e.g., word

order, case marking, agreement). If more than one cue is available in a given language, the most

frequent and hence reliable cue should have the largest impact on processing strategies

(MacWhinney & Bates, 1989). This hypothesis has been supported by data from English, German

and Russian (Arnon, Snider, Hofmeister, Jaeger, & Sag, 2006; Kempe & MacWhinney, 1999;

MacWhinney & Bates, 1989; Temperly & Gidea, 2010) and by data from an artificial language

learning study (Fedzechkina et al., 2012). As sentence word order freedom increases, its

informativity about the underlying syntactic structure might decrease (Fedzechkina, Jaeger, &

Newport, 2013; Kurumada & Jaeger, 2013) and therefore the impact of constituent length on

sentence word order might lessen as a result. Hence, English, Japanese and Korean speakers

generally shifted word order more often than Basque speakers because word order is a

significantly less reliable processing cue in Basque than in the other languages. This approach is

also broadly compatible with information-theoretic and inference-based accounts that link

production preferences to efficient communication (Gibson et al., 2013; Jaeger, 2010; for further

discussion and references see Jaeger & Tilly, 2011).

*The role of cross-linguistic differences on the impact of conceptual accessibility*

Yamashita and Chang (2001) argue that long phrases are conceptually more salient, and

speakers will tend to produce conceptually salient information early, to the extent that it does not

conflict with other word order constraints such as fixed word order. Specifically, they argue that

the freer the word order of a language, the stronger the effects of conceptual saliency and hence

the tendency to place longer elements before shorter ones. Our results do not support this

hypothesis. As discussed before, Basque shows a smaller degree of NP-shifting as a result of

phrase length than Japanese or Korean, even though it has a freer sentence word order. Thus,

conceptual saliency does not appear to be a critical factor in cross-linguistic word order

differences relative to length. Note that this should not be taken to mean that availability effects

do not play any role in OV languages. Rather, in our data dependency minimization strategies

seem to play a more important role during sentence production planning (e.g., there is plenty of

evidence showing availability effects in OV languages; Branigan et al., 2008; Ferreira & Yoshita,

2003, among others).

*Conclusion*

Our study sheds new light on the factors determining sentence word order preferences,

contributing to our understanding of how language processing is modulated by demands

interacting with variations in the linguistic properties of grammars. The *Performance Theory of*

*Order and Constituency* model (*PTOC*) (Hawkins, 1994, 2004) claims that minimising processing

domains makes language processing more efficient. The *PTOC* predicts a short-before-long

preference in VO languages and a long-before-short tendency in OV languages. The general

prediction regarding the different preferences in VO versus OV languages is met in our results.

Specifically, when the sentence word order produced by our participants departed from the

canonical one, long NPs preceded shorter ones. Additionally, the impact of length appeared to be

smaller in Basque, a language with freer word order than all previously studied ones, which might

indicate a possible trade-off between word order freedom and the impact of length on sentence

word order preferences.

Hawkins (1994, 2014) proposes a metric that measures the degree of parsing efficiency,

yielding precisely ranked predictions regarding sentence word order preferences in language

production. This metric predicts that long-before-short verb-final sentences will be maximally

efficient in Basque and thus will be preferred by speakers. These predictions were not fully

consistent with our results: when the object was long, verb-medial sequences were favoured over

their verb-final counterparts. We argue that verb agreement is the factor behind this preference

for shifted verb-medial word orders. In contrast to rigidly verb-final OV languages that lack verb

agreement, in heavily inflected flexible OV languages like Basque a fronted verb that carries

agreement morphology is a very reliable cue to sentence structure. In other words, differences in

the morpho-syntactic cues available in the language yield different patterns of sentence word

order preferences, although the principles behind them are common to all languages. All in all,

our study shows that cross-linguistic research is crucial for understanding how universal

processing demands interact with different types of grammatical specifications, yielding different

production patterns that emerge from a common source (echoing Hawkins, 1994, 2004).

**References**

Aldai, G. (2011). Wh-questions and SOV languages in Hawkins' (2004) theory: Evidence from Basque. *Linguistics*, *49–45,* 1079–1135. DOI: 10.1515/LING.2011.030

Arnold, J. E., Wasow, T., Asudeh, A., & Alrega, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language*, *55*, 55–70. DOI: 10.1016/j.jml.2004.03.006

Arnold, J. E., Wasow, T., Losongco, T., & Ginstrom, R. (2000). Heaviness vs. newness: the effects of structural complexity and discourse status on constituent ordering. *Language, 76*, 28–55. DOI: 10.2307/417392

Arnon, I., Snider, N., Hofmeister, P., Jaeger, T. F., & Sag, I. A. (2006). Cross-linguistic variation in a processing account: The case of multiple wh-questions. *Proceedings of 32 Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: BLS..

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. DOI: 10.1016/j.jml.2007.12.005

Bates, D., Maechler, M., & Bolker, B. (2011). l*me4: Linear mixed-effects models using S4 classes. R package (version 0.*999375-42). <http://CRAN.R-project.org/package=lme4>. Accessed 20.07.13.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions
to sentence formulation. *Psychological Review, 89*, 1–47. DOI: 10.1037/0033-
295X.89.1.1

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence
production. *Journal of Verbal Learning and Verbal Behavior*, *19*, 467–484.

Bock, J. K., & Levelt, J. W. M. (1994). Language production: Grammatical encoding. In M. A.
Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). Orlando, FL: Academic
Press.

Bock, J. K., & Warren, R. (1985). Conceptual accessibility and syntactic structure in sentence
formulation. *Cognition, 21*, 47–67.

Branigan, H., Pickering, M., & Tanaka, M. (2008). Contributions of animacy to grammatical
function assignment and word order during production. *Lingua, 118*, 172–189. DOI:
10.1016/j.lingua.2007.02.003

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G.
Boume, I. Kraemer, & J. Zwarts (Eds), *Cognitive foundations of interpretation* (pp. 69–
94). Amsterdam: Royal Netherlands Academy of Science.

Chang, F. (2009). Learning to order words: A connectionist model of heavy NP shift and
accessibility effects in Japanese and English. *Journal of Memory and Language, 61*,
374–397.

Choi, H. W. (2007). Length and order: a corpus study of Korean dative-accusative construction.
*Discourse and Cognition, 14*, 207–227. DOI:

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky , N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Collins, M. (2003). Head-driven statistical models for natural language parsing. *Computational
Linguistics, 29*, 589–638. DOI: 10.1162/089120103322753356

de Marneffe, M.C., Grimm, S., Arnon, I., Kirby, S., & Bresnan, J. (2012). A statistical model of the

grammatical choices in child production of dative sentences. *Language and Cognitive

Processes 27,* 25–61. DOI: 10.1080/01690965.2010.542651

de Rijk, R. P. G. (1969). Is Basque an SOV language? *Fontes Linguae Vasconum, 1–3*, 319–351

(Reprinted in De Rijk, R. P. G. (1998). *De Lingua Vasconum: Selected writings.*

*Supplements of the Anuario del Seminario de Filología Vasca "Julio de Urquijo", XLLIII*).

de Rijk, R. P. G. (2007). *Standard Basque, a progressive grammar.* Cambridge, MA: MIT Press.

de Smedt, K. (1994). Parallelism in incremental sentence generation. In G. Adriaes, & U. Hahn

(Eds.), *Parallelism in Natural Language Processing* (pp. 421–447). Norwood, NJ: Able.

Dennison, H. Y. (2008). Universal versus language-specific conceptual effects on shifted word-

order production in Korean: evidence from bilinguals. *Working Papers in Linguistics:

University of Hawaii at Manoa 39(2)*, 1–16.

Dryer, M. S. (1980). The positional tendencies of sentential noun phrases in universal grammar.

*Canadian Journal of Linguistics 25*, 123–195.

Dryer, M. S. (1992). The Greenbergian word order correlations. *Language, 68*, 81–138. DOI:

10.1353/lan.1992.0028

Dryer, M. S., & Haspelmath, M. (Eds.) (2013). *The World Atlas of Language Structures Online*.

Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info)

Eisner, J., & Smith, N. (2005). Parsing with soft and hard constraints on dependency length.

*Proceedings of the 9th International Workshop on Parsing Tachnology (IWPT 2005)*, 30–

41. DOI: 10.1007/978-90-481-9352-3_8

Elordieta, A. (2001). *Verb movement and constituent permutation in Basque*. Leiden: Landelijke

Onderzoekschool Taalwetenschap.

Erdocia, K., Laka, I., Mestres-Missé, 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

Fedzechkina, M., Jaeger, T. F., & Newport, E. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, 1–6. DOI: 10.1073/pnas.1215776109

Fedzechkina, M., Jaeger, T. F., & Newport, E. (2013).Communicative biases shape structures of newly acquired languages. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (*CogSci13*), 430–435. Austin, TX: Cognitive Science Society.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums*. Journal of Memory and Language, 46*, 57–84. DOI: 10.1006/jmla.2001.2797

Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language, 35*, 724–755. DOI: 10.1006/jmla.1996.0038

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*, 296–340. DOI: 10.1006/cogp.1999.0730

Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research, 32*, 669–692. DOI: 10.1023/A:1026146332132

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers, 35*, 116–124. DOI: 10.3758/BF03195503

Gennari, S. P., Jelena, M., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology, 65*, 141–176. DOI: 10.1016/j.cogpsych.2012.03.002

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*, 1–76.

DOI: 10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic

complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain* (pp.

95–126). Cambridge, MA: MIT Press.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A Noisy-Channel

Account of Cross-Linguistics Word Order Variation. *Psychological Science, 4*, 1079–

1088. DOI: 10.1177/0956797612463705

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological

encoding. *Cognition, 82*, B1–B14. DOI: 10.1016/S0010-0277(01)00138-X

Griffin, Z. M., & Bock, J. K. (2000). What the eyes say about speaking. *Psychological Science,

11*, 274–279. DOI: 10.1111/1467-9280.00255

Hakuta, K. (1981). Grammatical description versus configurational arrangement in language

acquisition: The case of relative clauses in Japanese. *Cognition, 9*, 197–236. DOI:

10.1016/0010-0277(81)90001-9

Hawkins, J. A. (1983). *Word order universals*. New York, NY: Academic Press.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge, UK:

Cambridge University Press. DOI: 10.1017/CBO9780511554285

Hawkins, J. A. (1998). Some issues in a performance theory of word order. In A. Siewierska (Ed.),

*Constituent Order in the Languages of Europe* (pp. 729–781). Berlin: Mouton de Gruyter.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford, UK: Oxford University

Press. DOI: 10.1093/acprof:oso/9780199252695.001.0001

Hawkins, J. A. (2009). An efficiency theory of complexity and related phenomena. In G. Sampson,

D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (pp. 158–175).

Cambridge, UK: Oxford University Press.

Hawkins, J. A. (2014). *Cross-linguistic variation and efficiency*. Oxford, UK: Oxford University

Press. DOI: 10.1093/acprof:oso/9780199664993.001.0001

Hinds, J. (1986). *Japanese.* London: Routledge.

Hoffman, C. (1999). Word order and the principle of "Early Immediate Constituents" (EIC).

*Journal of Quantitative Linguistics, 6*, 108–116. DOI: 10.1076/jqul.6.2.108.4133

Hualde, J. I., & Ortiz de Urbina, J. (2011). *A Grammar of Basque*. Berlin: Mouton de Gruyter.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and

towards logit mixed models. *Journal of Memory and Language, 59*, 434–446. DOI:

10.1016/j.jml.2007.11.007

Jaeger, T. F. (2010). Redundancy *and* reduction: Speakers manage syntactic information density.

*Cognitive Psychology*, *61*, 23–62. DOI:10.1016/j.cogpsych.2010.02.002.

Jaeger, T. F. (2011*).* Corpus-based research on language production information density and

reducible subject relatives. In E. M. Bender & J. E. Arnold (Eds.), *Language from a*

*cognitive perspective: Grammar, usage, and processing. Studies in honor of Tom Wasow*

(pp. 161– 197). Standford, CA: CSLI Publications.

Jaeger, T. F., & Norcliffe, E. (2009). The cross-linguistic study of sentence production. *Language*

*and Linguistic Compass, 3*, 866–887. DOI: 10.1111/j.1749-818x.2009.00147.x

Jaeger, T. F., & Tilly, H. (2011) Language Processing Complexity and Communicative Efficiency.

*WIRE: Cognitive Science, 2*, 323–325. DOI: 10.1002/wcs.126

Kempe, V., & MacWhinney, B. (1999). Processing of morphological and semantic cues in Russian

and German. *Language and Cognitive Processes*, *14*, 129–171. DOI:

10.1080/016909699386329

Kizach, J. (2012). Evidence for weight effects in Russian. *Russian Linguistics, 36*, 251–270. DOI:

10.1007/s11185-012-9096-0

Köhler, R. (1999). Syntactic structures: Properties and interrelations. *Journal of Quantitative Linguistics*, 6, 46–57. DOI: 10.1076/jqul.6.1.46.4137

Kondo, T., & Yamashita, H. (2011). Why speakers produce scrambled sentences: an analysis of a spoken language corpus in Japanese. In H. Yamashita, Y. Hirose, & J. L. Packard (Eds.), *Processing and Producing Head-Final Structures* (pp. 195–217). New York, NY: Springer. DOI: 10.1007/978-90-481-9213-7

Konopka, A. E. & Brown-Schmidt, S. (2014). Message encoding. In V. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *The Oxford Handbook of language production* (pp. 3–20). New York: Oxford University Press.

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology, 73,* 1–40. DOI: 10.1016/j.cogpsych.2014.04.001

Kurumada, C., & Jaeger, T. F. (2013). Communicatively efficient language production and case-marker omission in Japanese. *Proceedings of the 35th Annual Meeting of Cognitive Science Society (CogSci14),* 858–863. Austin, TX: Cognitive Science Society.

Laka, I. (1993). The Structure of Inflection: A Case Study in X° Syntax. In J. I. Hualde & J. Ortiz de Urbina (Eds.), *Generative Studies in Basque Linguistics* (pp. 21–70). Amsterdam/Philadelpia: John Benjamins.

Laka, I. (1994) *On the Syntax of Negation*. Outstanding Dissertations in Linguistics Series. New York-London: Garland Publishing Co.

Laka, I. (1996). *A Brief Grammar of Euskara, the Basque Language*. University of the Basque Country, Office for the vicerector for the Basque Language. Retrieved from http://www.ehu.es/grammar

Levelt, W. J. M., & Maassen, B. (1981). Lexical search and order of mention in sentence production. In W. Klein & W. J. M. Levelt (Eds.), *Crossing the boundaries in linguistics* (pp. 221–252). Dordrecht, The Netherlands: D. Reidel. DOI: 10.1007/978-94-009-8453-

0_12

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic

choices. *Psychonomic Bulletin and Review, 9*, 550–557. DOI: 10.3758/BF03196312

Lohse, B., Hawkins, J., & Wasow, T. (2004). Domain minimization in English verb-particle

construction. *Language*, *80*, 238–261. DOI: 10.1353/lan.2004.0089

MacDonald, M. C. (2013). How language production shapes language from and comprehension.

*Frontiers in Psychology, 4*, 1–16. DOI: 10.3389/fpsyg.2013.00226

MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of*

*language acquisition* (pp. 249–308). HIllsdale, NJ: Lawrence Erlbaum.

MacWhinney, B., & Bates, E. (Eds.) (1989). *The crosslinguistic study of sentence processing*.

New York, NY: Cambridge University Press.

Marblestone, K. L. (2007). *Semantic and syntactic effects on double prepositional phrase*

*ordering across the lifespan.* Unpublished doctoral dissertation. University of Southern

California, Los Angeles.

Maurits, L, Perfors, A., & Navarro, D. (2010). Why are some word orders more common than

others? A uniform information density account. *Advances in Neural Information*

*Processing Systems, 23*, 1585–1593.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R.

Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 2) (pp. 419–

491). New York, NY: Wiley.

Pablos, L. (2011). Rich agreement in Basque: Evidence for pre-verbal structure building. In H.

Yamashita, Y. Hirose, & J. L. Packard (Eds.), *Processing and producing head-final*

*structures* (pp. 3–21). New York, NY: Springer. DOI: 10.1007/978-90-481-9213-7

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in

    language production: a cross-linguistic study in English and Spanish. *Journal of Memory*

    *and Language, 42*, 168–82. DOI: 10.1006/jmla.1999.2668

Primus, B. (1999). *Case and thematic roles: Ergative, accusative and active.* Tübingen:

    Niemeyer. DOI: 10.1515/9783110912463

R Development Core Team (2012). R: A Language and Environment for Statistical Computing. R

    Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org>.

Saito, M. (2004). Japanese scrambling in a comparative perspective. In D. Adger, D. de Cat, & G.

    Tsoulas (Eds.), *Peripheries: Syntactic edges and their effects* (pp. 143–163). Dordrecht:

    Kluwer Academic Publishers.

Sleator, D., & Temperley, D. (1991). Parsing English with a link grammar. Technical report CMY-

    CS-91-196. Pittsburgh, PA: Carnegie Mellon University, School of Computer Science.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production.

    *Cognition, 73*, 205–246. DOI: 10.1016/S0010-0277(99)00053-0

Smith, M., & Wheeldon, L. (2004). Horizontal information flow in spoken sentence production.

    *Journal of Experimental Psychology: Learning Memory and Cognition, 30*, 675–686. DOI:

    10.1037/0278-7393.30.3.675

Sohn, H. (1994). *Korean*. London/New York: Routledge.

Stallings, L. M., & MacDonald, M. C. (2011). It's not just the "heavy NP": Relative phrase length

    modulates the production of heavy-NP shift. *Journal of Psycholinguistics, 40*, 177–187.

    DOI: 10.1007/s10936-010-9163-x

Stallings, L. M., MacDonald, M. C., & O'Seaghdha, P. G. (1998). Phrasal ordering constraints in

    sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of*

    *Memory and Language, 39*, 392–417. DOI: 10.1006/jmla.1998.2586

Szmrecsányi, B. (2004). On operationalizing syntactic complexity. *Journées Internationals d'Analyse statistique des Données Textualles, 7*, 1031–1038.

Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language, 65,* 318–330. DOI: 10.1016/j.jml.2011.04.009

Tanaka, M. N., Branigan, H. P., & Pickering, M. J. (2011). The production of head-initial and head-final languages. In H. Yamashita, Y. Hirose, & J. L. Packard (Eds.), *Processing and producing head-final structures* (pp. 113–129). New York, NY: Springer. DOI: 10.1007/978-90-481-9213-7

Temperley, D. (2007). Minimization of dependency length in written English. *Cognition, 105*, 300–333. DOI: 10.1016/j.cognition.2006.09.011

Temperley, D. (2008). Dependency-Length Minimization in Natural and Artificial Languages. *Journal of Quantitative Linguistics*, 15, 256–282. DOI: 10.1080/09296170802159512

Temperly, D., & Gidea, D. (2010). Do Grammars Minimize Dependency Length? *Cognitive Science, 34*, 286–310. DOI: 10.1111/j.1551-6709.2009.01073.x

Uszkoreit, H. (1987). *Word order and constraint structure in German*. Stanford, CA: CSLI Publications.

Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, L., Oepen, S., & Skut, W. (1998). Studien zur performanzorientierten Linguistik: Aspekte der Relativsatzextraposition in Deutschen. [Studies in performance-oriented Linguistics: Aspects of German Relative Clause Extraposition]. *Kognitionswissenschaft, 7*, 129–133.

Wasow, T. (1997a). End-weight from the speaker's perspective. *Journal of Psycholinguistics Research, 26*, 347–362. DOI: 10.1023/A:1025080709112

Wasow, T. (1997b). Remarks on grammatical weight. *Language Variation and Change, 9*, 81-105. DOI: 10.1017/S0954394500001800

Wasow, T. (2002). *Postverbal behavior*. Stanford, CA: CSLI Publications.

Wasow, T. (2013). The appeal of the PDC program. *Frontiers in Psychology, 4*, 1–3. DOI:

    10.3389/fpsyg.2013.00236

Wiechmann, D. & Lohmann, A. (2013). Domain minimization and beyond: Modelling PP ordering.

    *Language Variation and Change, 25*, 65–88.

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

**Footnotes**

[1] Definitions of length based on either number of words or of phrasal nodes have been proved to

be strongly correlated (Szmrecsányi, 2004; Wasow, 2002).

[2] The list of abbreviations used in the glosses is the following: Aux: Auxiliary; Det: Determinant;

IO: Indirect Object, which is marked with Dat(ive case); O: Direct Object, which is zero marked; S:

Subject, which in the sentences used in this paper are all Erg(ative) case-marked.

[3] Note that although taking German to be a VO language is highly controversial, here we are

assuming Hawkins' classification (1994: 393-400; 2014: 177). However, in the *World Atlas of*

*Language Structures Online* (http//wals.info), German is classified as both SOV and SVO. We

thank an anonymous reviewer for pointing this out.

[4] Arguably, Hawkins' model defines efficiency from the perspective of the hearer rather than of the

speaker, and makes a number of assumptions about the interaction of comprehension and

production that require further research (for a more detailed discussion, see Hawkins, 1998).

[5] Auxiliaries immediately follow the verb except in negative and emphatic sentences where the

auxiliary precedes the lexical verb (Laka 1994, 1996). The semantic information carried by the

lexical verb is often not sufficient to determine argument structure: for instance, the verb *hil* can

mean either "die" when it is construed with one argument, or it can mean "kill" when constructed

with two arguments:

(i) katua hil da          (ii) txakurrak katua hil du

Cat-Det dead is            dog-Det-Erg cat-det kill has

'The cat is dead'          'the dog killed the cat'

[6] In the case of ditransitive sentences, the best model included the Position of the constituents in

the Screen, Length and their interaction as fixed effects and a by-participant random slope for

Length. In this model the effect of long-IO did not replicate ($p < .5$). Also, using only the data for

which we have most data points, namely verb-final cases, the effect of long-O but not of long-IO

replicated in the ditransitive sentences ($p < .01$).

*Table 1. MiD and MaOP metrics for the transitive structures in the Long-O condition.*

| Order of constituents | Example |
| --- | --- |
| **O-S-V** | [$_O$ Geltokian trenaren zain egon den gaztea] [$_S$ andreek] [$_V$ agurtu dute] |
| | [$_O$ Station-in train-of wait be AUX-that young person-Det] [$_S$ women-Erg] [$_V$ greeted AUX$_{O.S}$] |
| | The women greeted the young man that had been waiting for the train at the station |
| | ***MiD: 83.4%          MaOP: lower*** |
| | S': 2/2 = 100% |
| | Number of words: {andreek, agurtu}. ICs: 2, {NP, VP} |
| | VP: 2/3 = 66.7% |
| | Number of words: {gaztea, andreek, agurtu}. ICs: 2, {NP, VP} |
| **O-V-S** | [$_O$ Geltokian trenaren zain egon den gaztea] [$_V$ agurtu dute] [$_S$ andreek] |
| | [$_O$ Station-in train-of wait be AUX-that young person-Det] [$_V$ greeted AUX$_{O.S}$] [$_S$ women-Erg] |
| | The women greeted the young man that had been waiting for the train at the station |
| | ***MiD: 83.4%          MaOP: lowest*** |
| | S': 2/3 = 66.7% |
| | Number of words: {agurtu, dute, andreek}. ICs: 2, {NP, VP} |
| | VP: 2/2 = 100% |
| | Number of words: {gaztea, agurtu}. ICs: 2, {NP, VP} |

| | |
|---|---|
| **S-O-V** | [s Andreek] [o geltokian trenaren zain egon den gaztea] [v agurtu dute] |
| | [s Women-Erg] [o station-in train-of wait be AUX-that young person-Det] [v greeted AUX$_{O.S}$] |
| | The women greeted the young man that had been waiting for the train at the station |
| | *MiD: 62.5%*      **MaOP: high** |
| | S': 2/8 = 25% |
| | Number of words: {andreek, trenaren, zain, egon, den, gaztea, agurtu}. ICs: 2, {NP, VP} |
| | VP: 2/2 = 100% |
| | Number of words: {gaztea, agurtu}. ICs: 2, {NP, VP} |
| **S-V-O** | [s Andreek] [v agurtu dute] [o geltokian trenaren zain egon den gaztea] |
| | [s Women-Erg] [v greeted AUX$_{O.S}$] [o station-in train-of wait be AUX-that young person-Det] |
| | The women greeted the young man that had been waiting for the train at the station |
| | *MiD: 62.5%*      **MaOP: high** |
| | S': 2/2 = 100% |
| | Number of words: {andreek, agurtu}. ICs: 2, {NP, VP} |
| | VP: 2/8 = 25% |
| | Number of words: {agurtu, geltokian, trenaren, zain, egon, den, gaztea }. ICs: 2, {NP, VP} |

List of abbreviations: Aux: Auxiliary; Erg: Ergative; ICs: Immediate Constituents; O: Direct Object;

S': Sentence; S: Subject; V: Verb.

*Table 2. MiD and MaoP metrics for the ditransitive structures in the Long-O condition.*

| Order of constituents | Example |
|---|---|
| **O-IO-S-V** | [O Gurasoek gogoko duten pastela] [IO irakasleari] [S amak] [V ekarri dio] |
| | [O Parents like AUX-that cake-Det] [IO teacher-Dat] [S mother-Erg] [V took AUX$_{O.IO.S}$] |
| | The mother brought the teacher the cake that parents like |
| | ***MiD: 87.5%***                    ***MaOP: lowest*** |
| | S': 2/2 = 100% |
| | Number of words: {amak, ekarri }. ICs: 2, {NP, VP} |
| | VP: 3/4 = 75% |
| | Number of words: {*pastela, irakasleari, amak, ekarri* }. ICs: 3, {NP, NP, VP} |
| **S-O-IO-V** | [S Amak] [O gurasoek gogoko duten pastela] [IO irakasleari] [V ekarri dio] |
| | [S Mother-Erg] [O parents like AUX-that cake-Det] [IO teacher-Dat] [V took AUX$_{O.IO.S}$] |
| | The mother brought the teacher the cake that parents like |
| | ***MiD: 64.3%***          ***MaOP: lower*** |
| | S':  2/7 = 28.6% |
| | Number of words: {amak, irakasleari, gurasoek, gogoko, duten, pastela, ekarri}. ICs: 2, {NP, VP} |
| | VP: 3/3 = 100% |
| | Number of words: {pastela, irakasleari, ekarri }. ICs: 3, {NP, NP, VP} |

| | |
|---|---|
| **O-V-S-IO** | [O Gurasoek gogoko duten pastela] [V ekarri dio] [S amak] [IO irakasleari]<br><br>[O Parents like AUX-that cake-Det] [V took AUX$_{O.IO.S}$] [S mother-Erg] [IO teacher-Dat]<br><br>The mother brought the teacher the cake that parents like<br><br>**MiD: 63.4%**             **MaOP: lowest**<br><br>S': 2/3 = 66.7%<br><br>Number of words: {*ekarri, dio, amak*}. ICs: 2, {NP, VP}<br><br>VP: 3/5 = 60%<br><br>Number of words: {*pastela, ekarri, dio, amak, irakasleari* }. ICs: 3, {NP, NP, VP} |
| **S-O-V-IO** | [S Amak] [O gurasoek gogoko duten pastela] [V ekarri dio] [IO irakasleari]<br><br>[S Mother-Erg] [O parents like AUX-that cake-Det] [V took AUX$_{O.IO.S}$] [IO teacher-Dat]<br><br>The mother brought the teacher the cake that parents like<br><br>**MiD: 54,2%**          **MaOP: lower**<br><br>S': 2/6 = 33.3%<br><br>Number of words: {*amak, gurasoek, gogoko, duten, pastela, ekarri*}. ICs: 2, {NP, VP}<br><br>VP: 3/4 = 75%<br><br>Number of words: {*pastela, ekarri, dio, irakasleari* }. ICs: 3, {NP, NP, VP} |
| **S-IO-O-V** | [S Amak] [IO irakasleari] [O gurasoek gogoko duten pastela] [V ekarri |

---

dio]

[$_S$ Mother-Erg] [$_{IO}$ teacher-Dat] [$_O$ parents like AUX-that cake-Det] [$_V$ took AUX$_{O.IO.S}$]

The mother brought the teacher the cake that parents like

***MiD: 39.3%***          ***MaOP: high***

S':  2/7  = 28.6%

Number of words: {amak, irakasleari, gurasoek, gogoko, duten, pastela, ekarri}. ICs: 2, {NP, VP}

VP: 3/6 = 50%

Number of words: {irakasleari, gurasoek, gogoko, duten, pastela, ekarri}. ICs: 3, {NP, NP, VP}

---

List of abbreviations: Aux: Auxiliary; Dat: Dative; Erg: Ergative; IO: Indirect Object; ICs: Immediate Constituents; O: Direct Object; S': Sentence; S: Subject; V: Verb.

*Table 3. Ranking of sentence word orders as predicted by the PTOC model for Basque.*

The predicted ranking of sentence word orders for the type of sentences used in our experiment are shown in decreasing order of preference from top to bottom in each column, according to MiD (efficiency ratios are given in parenthesis), to MaOP and to the interaction of MiD and MaOP.

| MiD-predicted | MaOP-predicted | MID and MaOP-predicted |
|---|---|---|
| *Transitive sentences: Long-O condition* | | |
| O-S-V and O-V-S (both 83.4%) | S-O-V and S-V-O | O-S-V |
| S-O-V and S-V-O (both 62.5%) | O-S-V | O-V-S |
|  | O-V-S | S-O-V and S-V-O |
| *Ditransitive sentences: Long-O condition* | | |
| O-IO-S-V (87.5%) | S-IO-O-V | O-IO-S-V |
| S-O-IO-V and O-V-S-IO (64.3%, 63.4%) | S-O-IO-V | S-O-IO-V |
| S-O-V-IO (54.2%) | S-O-V-IO | O-V-S-IO |
| S-IO-O-V (39.3%) | O-IO-S-V and O-V-S-IO | S-O-V-IO |
|  |  | S-IO-O-V |

*Table 4. Language history and self-assessed proficiency mean ratings (standard deviations in brackets) reported by the participants.*

| | Infancy (0-3 years) | Primary school (4-12 years) | Secondary school (12-18 years) | Adulthood (after 18) |
|---|---|---|---|---|
| *Self-assessed language use* | | | | |
| Home | 1.25 (.44) | 1.17 (.48) | 1.21 (.51) | 1.21 (.51) |
| School | - | 1.50 (.66) | 1.71 (.55) | 1.71 (.95) |
| Other | - | 1.71 (.55) | 1.83 (.48) | 2.04 (.46) |

| | *Basque* | *Spanish* |
|---|---|---|
| *Self-rated proficiency level* | | |
| Listening | 6.83 (.38) | 6.00 (.72) |
| Speaking | 6.79 (.41) | 5.00 (.88) |
| Reading | 6.71 (.46) | 5.67 (.86) |
| Writing | 6.42 (.72) | 5.17 (.76) |
| MEAN | 6.69 (.49) | 5.46 (.80) |

*Self-assessed language use rating scores on a 7-point scale*: 1: Only Basque; 2: Mostly Basque, a few times Spanish; 3: Mostly Basque but Spanish at least 25% of the time; 4: Basque and

Spanish equally; 5: Mostly Spanish but Basque at least 25% of the time; 6: Mostly Spanish, a few

times Basque; 7: Only Spanish.

*Self-rated proficiency level scores on a 7-point scale*: 7: Native-like level; 6: High level; 5:

Medium-high level; 4: Medium level; 3: Medium-low level; 2: Low level; 1: No knowledge.

*Table 5.* Examples of the three different types of length experimental conditions used for transitive sentences in the experiment.

| Phrase length condition | Example |
| --- | --- |
| *Transitive sentences* | |
| All-Short | Andreek gaztea agurtu dute |
| | Woman-Det-S young person-Det.O greeted AUX$_{O.S}$ |
| | The women greeted the young man |
| Long-S | Geltokian trenaren zain egon diren andreek gaztea agurtu dute |
| | Station-in train-of wait be AUX-that woman-Det-S young person-Det.O greeted AUX$_{O.S}$ |
| | The women that had been waiting for the train at the station greeted the young man |
| Long-O | Andreek geltokian trenaren zain egon den gaztea agurtu dute |
| | Woman-Det-S station-in train-of wait be AUX-that young person-Det.O greet AUX$_{O.S}$ |
| | The women greeted the young man that had been waiting for the train at the station |

*Table 6.* Examples of the three different types of length experimental conditions used for ditransitive sentences in the experiment.

| Phrase length condition | Example |
| --- | --- |
| *Ditransitive sentences* | |
| All-Short | Amak irakasleari pastela eraman dio |
| | Mother-Det-S teacher-Det-IO cake-Det.O took $AUX_{O.IO.S}$ |
| | The mother brought the cake to the teacher |
| Long-IO | Amak gurasoek gogoko duten irakasleari pastela eraman dio |
| | Mother-Det-S parents-Det-S like AUX-that teacher-Det-IO cake-Det.O |
| | took $AUX_{O.IO.S}$ |
| | The mother brought the cake to the teacher that parents like |
| Long-O | Amak irakasleari gurasoek gogoko duten pastela eraman dio |
| | Mother-Det-S teacher-Det-IO parents-Det-S like AUX-that cake-Det.O |
| | took $AUX_{O.IO.S}$ |
| | The mother brought the cake that parents like to the teacher |

Table 7. Raw data and percentages (in brackets) of shifted and non-shifted structures relative to verb position in transitive sentences.

| Type of Response | | | Length Condition | | |
|---|---|---|---|---|---|
| Constituent order | Verb Position | Structure order | All-short | Long-S | Long-O |
| **S-O** | Final | S-O-V | 185 (96.4%) | 162 (88.5%) | 148 (78.7%) |
| **(non-shifted)** | Medial | S-V-O | 0 (0%) | 11 (6.0%) | 3 (1.6%) |
| *Total S-O* | | | *185 (96.4%)* | *173 (94.5%)* | *151 (80.3%)* |
| **O-S** | Final | O-S-V | 2 (1.0%) | 4 (2.2%) | 9 (4.8%) |
| **(shifted)** | Medial | O-V-S | 5 (2.6%) | 6 (3.3%) | 28 (14.9%) |
| *Total-O-S* | | | *7 (3.6%)* | *10 (5.5%)* | *37 (19.7%)* |
| **Miscellaneous** | | | 0 | 9 | 4 |

*Table 8.* Raw data and percentages (in brackets) of shifted and non-shifted structures relative to verb position in ditransitive sentences (verb-medial and verb-final orders with less than 5% in all of the conditions occurrences subsumed under *Others*).

| Type of Response | | | Length Condition | | |
|---|---|---|---|---|---|
| Constituent order | Verb Position | Structure order | All-short | Long-IO | Long-O |
| **IO-O** | Final | S-IO-O-V | 130 (54.9%) | 117 (50.2%) | 87 (37.3%) |
| **(non-shifted)** | | Others | 0 (0%) | 5 (2.2%) | 0 (0%) |
| | Medial | IO-O-V-S | 1 (.4%) | 17 (7.3%) | 0 (0%) |
| | | Others | 6 (2.5%) | 17 (7.3%) | 6 (2.6%) |
| | *Total-Final* | | 130 (54.9%) | 122 (52.4%) | 87 (37.3%) |
| | *Total-Medial* | | 7 (2.9%) | 35 (14.6%) | 6 (2.6%) |
| *Total IO-O* | | | *137 (57.8%)* | *157 (67%)* | *94 (39.9%)* |
| **O-IO** | Final | Others | 7 (3%) | 3 (1.3%) | 11 (4.7%) |
| **(shifted)** | Medial | S-O-V-IO | 88 (37.1%) | 71 (30.5%) | 89 (38.2%) |
| | | O-V-S-IO | 4 (1.7%) | 0 (0%) | 36 (15.5%) |
| | | Others | 0 (0.4%) | 3 (1.2%) | 4 (1.7%) |
| | *Total-Final* | | 7 (3%) | 3 (1.3%) | 11 (4.7%) |
| | *Total-Medial* | | 92 (39.2%) | 74 (31.7%) | 129 (55.4%) |
| *Total O-IO* | | | *99 (42.2%)* | *77 (33%)* | *140 (60.1%)* |
| **Miscellaneous** | | | 4 | 4 | 6 |

*Table 9.* Logit mixed model analyses for transitive and ditransitive sentences

|  | Estimate | Std. Error | z-value | p-value |
|---|---|---|---|---|
| *Transitive sentences*: | | | | |
| Main model: N = 563 | | | | |
| Formula: "O-S" ~ 1 + Length + (1 \| subject) + (1 \| item) | | | | |
| Intercept | -4.147 | .568 | -7.301 | <.001 |
| Length: Long-S | .590 | .587 | 1.004 | .316 |
| Length: Long-O | 2.350 | .395 | 5.948 | <.001 |
| *Ditransitive sentences* | | | | |
| Main model: N=703 | | | | |
| Formula: "O-IO" ~ 1 + Length +  (1 + Length \| subject) + (1 \| item) | | | | |
| Intercept | -.278 | .321 | -.866 | <.386 |
| Length: Long-IO | -.624 | .317 | -1.970 | .049 |
| Length: Long-O | 1.417 | .525 | 2.700 | .006 |
| Verb position: N= 703 | | | | |
| Formula: "Medial" ~ 1 + Length + (1 + Length \| subject) + (1 \| item) | | | | |
| Intercept | -.484 | .351 | -1.380 | .168 |
| Length: Long-IO | .178 | .298 | .597 | .551 |
| Length: Long-O | .909 | .418 | 2.178 | .029 |

**Appendix 1. Experimental materials. Transitive sentences.**

List of transitive sentences used in the experiment. All sentences are presented in canonical [SOV] word order. In the experiment the order of the noun phrases was counterbalanced and the verb appeared in the upper left-right corner. English translations are presented in italics.

1. [Altuak eta oso azkarrak diren] ertzainek [altua eta oso azkarra den] bahitzailea atxilotu dute

*The policemen [that are very fast and smart] arrested the kidnapper [that is very fast and smart]*

2. [Gurasoek ondo hezi dituzten] mutikoek [gurasoek ondo hezi duten] auzokidea zigortu dute

*The little boys [that have been well brought up by their parents] punished the neighbour [that has been well brought up by their parents]*

3. [Jaietan beti mozkortzen diren] koadrilakoek [jaietan beti mozkortzen diren] laguna defendatu dute

*The gang [that always gets drunk at parties] protected the member [that always gets drunk at parties]*

4. [Jatorrak eta esker onekoak diren] modeloek [jatorra eta esker onekoa den] argazkilaria txalotu dute

*The models [that are nice and grateful] applauded the photographer [that is nice and grateful]*

5. [Jenio bizia duten] ikaskideek [jenio bizia duen] mutila jipoitu dute

*The classmates [that have a bad temper] beat up the boy [that has a bad temper]*

6. [Unibertsitatearen alde lan egiten duten] ikasleek [unibertsitatearen alde lan egiten duen] errektorea besarkatu dute

*The students [that work on behalf of the university] hugged the provost [that works on behalf of the university]*

7. [Zakarrak eta zinikoak diren] gaixoek [zakarra eta zinikoa den] medikua kritikatu dute

*The sick people [that are mean and cynical] criticised the doctor [that is mean and cynical]*

8. [Saria jaso duten] idazkariek [saria jaso duen] enpresaria goraipatu dute

*The secretaries [that have received the prize] praised the entrepreneur [that has received the prize].*

9. [Gure amak aspalditik ezagutzen dituen] poliziek [gure amak aspalditik ezagutzen duen] lapurra harrapatu dute

*The policemen [that our mother has known for a long time] caught the thief [that our mother has known for a long time]*

10. [Frontoira maiz joaten diren] zaletuek [frontoira maiz joaten den] pilotaria ikusi dute

*The fans [that often go to the jai-alai] saw the player [that often goes to the jai-alai]*

11. [Medikazioa sakelean gordetzen duten] agureek [medikazioa sakelean gordetzen duen] erizaina beldurtu dute

*The old men [that hide the medication in the pocket] scared the nurse [that hides the medication in the pocket]*

12. [Jendartean pozik egoten diren] kazetariek [jendartean pozik egoten den] margolaria elkarrizketatu dute

*The journalists [that are happy among people] interviewed the painter [that is happy among people]*

13. [Etengabe kolasean dabiltzan] txakurrek [etengabe jolasean dabilen] haurra poztu dute

*The dogs [that are continuously playing] cheered up the child [that is continuously playing]*

14. [Inoiz egin den kontraturik garestiena duten] jokalariek [inoiz egin den kontraturik garestiena duen] atezaina baztertu dute

*The players [that have the most expensive contract ever signed] excluded the goalkeeper [that has the most expensive contract ever signed]*

15. [Torneoan lehenengoz parte hartu duten] boxeolariek [torneoan lehenengoz parte hartu duen] epailea errespetatu dute

*The boxers [that have taken part in the tournament for the first time] respected the referee [that has taken part in the tournament for the first time]*

16. [Geltokian trenaren zain egon diren] andreek [geltokian trenaren zain egon den] gaztea agurtu dute

*The women [that have been waiting for the train at the station] greeted the young man [that has been waiting for the train at the station]*

17. [Egunero goiz-goizetik lanean diharduten] okinek [egunero goiz-goizetik lanean diharduen] nekazaria gonbidatu dute

*The bakers [that work every day since very early] invited the farmer [that works every day since very early in the morning]*

18. [Asanbladan denon aurrean hitz egin duten] langileek [asanbladan denon aurrean hitz egin duen] politikaria aipatu dute

*The workers [that have spoken in front of everyone at the meeting] mentioned the politician [that has spoken in front of everyone at the meeting]*

19. [Poltsikoak diruz beteta dituzten] alargunek [poltsikoak diruz beteta dituen] abokatua kontratatu dute

*The widows [that have the pockets full of money] hired the lawyer [that has the pockets full of money]*

20. [Samurki eta poliki kantatzen duten] ilobek [samurki eta poliki kantatzen duen] amona hunkitu dute

*The grandchildren [that sing slowly and tenderly] moved the grandmother [that sings slowly and tenderly]*

21. [Atzera eta aurrera ibiltzen diren] umeek [atzera eta aurrera ibiltzen den] begiralea harritu dute

*The children [that walk back and forth] surprised the instructor [that walks back and forth]*

22. [Laster lanik gabe geratuko diren] erizainek [laster lanik gabe geratuko den] gaixoa sendatu dute

*The nurses [that will be out of their jobs soon] treated the ill man [that will be out of his job soon]*

23. [Pelikula berriaren gidoia gustatzen zaien] aktoreek [pelikula berriaren gidoia gustatzen zaion] zuzendaria babestu dute

*The actors [that like the script of the new film] protected the director [that likes the script of the new film]*

24. [Oso zabarki jantzita ibiltzen diren] saltzaileek [oso zabarki jantzita ibiltzen den] etxekoandrea nazkatu dute

*The sellers [that are usually very poorly dressed] disgusted the housewife [that is usually very poorly dressed]*

**Appendix 2. Experimental sentences. Ditransitive sentences**

List of ditransitive sentences used in the experiment. All sentences are written in canonical [SIOOV] word order. In the experiment the order of the noun phrases was counterbalanced and the verb appeared in the upper left-right corner. English translations are presented in italics.

1. Edurnek [hainbat sari irabazi dituen] idazleari [hainbat sari irabazi dituen] tarta ekarri dio

*Edurne brought the cake [that won so many prizes] to the writer [that won so many prizes]*

2. Amak [gurasoek gogoko duten] irakasleari [gurasoek gogoko duten] pastela eraman dio

*The mother brought the pie [that parents like] to the teacher [that parents like]*

3. Amaiak [mahaitik urrunegi zegoen] emakumeari [mahaitik urrunegi zegoen] poltsa eman dio

*Amaia gave the bag [that was too far away from the table] to the woman [that was too far away from the table]*

4. Lehendakariak [hiriaren babesa zeukan] alkateari [hiriaren babesa zeukan] plana erakutsi dio

*The president showed the plan [that had the support of the city] to the mayor [that had the support of the city]*

5. Mirenek [ederra iruditzen zitzaion] dendariari [ederra iruditzen zitzaion] bertsoa kantatu dio

*Miren sang the verse [that seemed beautiful to her] to the shop assistant [that seemed beautiful to her]*

6. Aurkezleak [telebistan agertu den] zientzialariari [telebistan agertu den] notizia kontatu dio

*The host told the news [that were on television] to the scientist [that was on television]*

7. Itziarrek [herriko guztiek ezagutzen duten] kirolariari [herriko guztiek ezagutzen duten] museoa erakutsi dio

*Itziar showed the museum [that everybody in town knows] to the athlete [that everybody in town knows]*

8. Agureak [maite ez zuen] senideari [maite ez zuen] baserria utzi dio

*The old man left the farm [that he did not love] to the relative [that he did not love]*

9. Sarak [polita eta oso dotorea den] kantariari [polita eta oso dotorea den] argazkia egin dio

*Sara took a photo [that looks nice and elegant] of the singer [that looks nice and elegant]*

10. Musikariak [izugarri hunkigarria den] abeslariari [izugarri hunkigarria den] balada idatzi dio

*The musician wrote a ballad [that is very moving] to the singer [that is very moving]*

11. Kepak [oso aipatua den] marrazkilariari [oso aipatua den] komikia bidali dio

*Kepa sent the comic [that is very well known] to the painter [that is very well known]*

12. Argitaratzaileak [iaz arrakasta handia izan zuen] itzultzaileari [iaz arrakasta handia izan zuen] liburua eskaini dio

*The editor offered the book [that had a great success last year] to the translator [that had a great success last year]*

13. Mikelek [interesgarria ematen duen] adiskideari [interesgarria ematen duen] aldizkaria apurtu dio

*Mikel tore up the magazine [that looks interesting] of/to [his] friend [that looks interesting]*

14. Ahizpak [denoi asko gustatzen zaigun] emakumeari [denoi asko gustatzen zaigun] eleberria gomendatu dio

*The sister recommended a book [that we all like a lot] to the woman [that we all like a lot]*

15. Kirmenek [Eibarren oso ezaguna den] gizonari [Eibarren oso ezaguna den] garagadoa atera dio

*Kirmen bought a beer [that is very popular in Eibar] to the man [that is very popular in Eibar]*

16. Argazkilariak [egun guztietan ondoan izan duen] laguntzaileari [egun guztietan ondoan izan duen] kamera kendu dio

*The photographer took the camera [that he always has next to him] away from the assistant [that he always has next to him]*

17. Aitorrek [mundu osoan izugarri ospetsua den] pailazoari [mundu osoan izugarri ospetsua den] jokoa saldu dio

*Aitor sold the game [that is very famous all over the world] to the clown [that is very famous all over the world]*

18. Detektibeak [egun osoan zelatatu duen] harakinari [egun osoan zelatatu duen] furgoneta lapurtu dio

*The detective stole the van [that he had been spying on the whole day long] to the butcher [that he had been spying on the whole day long]*

19. Ainhoak [aspalditik asko kezkatzen duen] alabari [aspalditik asko kezkatzen duen] arazoa ezagutarazi dio

*Ainhoa explained the problem [that worries her since long ago] to the daughter [that worries her since long ago]*

20. Diputatuak [hainbat egunetan galduta ibili den] etorkinari [hainbat egunetan galduta ibili den] txakurra oparitu dio

*The deputy gave the dog [that had been lost for so many days] to the immigrant [that had been lost for  so many days]*

21. Asierrek [lehengo egunean dendan ikusi zuen] neskari [lehengo egunean dendan ikusi zuen] eraztuna erosi dio

*Asier bought the ring [that he saw at the store the other day] to the girl [that he saw at the store the other day]*

22. Funtzionarioak [etxean aurkitu izan duen] jabeari [etxean aurkitu izan duen] baimena izenpetu dio

*The civil servant signed the license [that he found at home] to the homeowner [that he found at home]*

23. Iratik [oso jende gutxik entzun duen] taldeari [oso jende gutxik entzun duen] abestia eskatu dio

*Irati asked the band [that so few people had listened to] for the song [that so few people had listened to]*

24. Zuzendariak [atzera bota zuten] ordezkariari [atzera bota zuten] proposamena itzuli dio

*The director returned the proposal [that they dismissed] to the substitute [that they dismissed]*

25. Zuriñek [pertsona guztien gustukoa den] gizonari [pertsona guztien gustukoa den] alkandora urratu dio

*Zuriñe tore the shirt [that everybody likes] of/to the man [that everybody likes]*

26. Presidenteak [enpresan lagun asko dituen] akziodunari [enpresan lagun asko dituen] bekaduna aurkeztu dio

*The president introduced the intern [that has a lot of friends at the firm] to the stockholder [that has a lot of friends at the firm]*

27. Jonek [gaizki-ulertua sortu duen] epaileari [gaizki-ulertua sortu duen] iruzkina zuzendu dio

*Jone corrected the remark [that caused the misunderstanding] to the judge [that caused the misunderstanding]*

28. Entrenatzaileak [garestia baina oso ona den] futbolariari [garestia baina oso ona den] baloia sinatu dio

*The manager signed the ball [that is expensive but very good] to the football player [that is expensive but very good]*

29. Josunek [neke handiz aurkitu duen] lankideari [neke handiz aurkitu duen] eserlekua gorde dio

*Josune saved the seat [that was very hard for her to find] for the colleague [that was very hard for her to find]*

30. Boluntarioak [oso kritikatua izan den] herritarrari [oso kritikatua izan den] erabakia azaldu dio

*The volunteer explained the decision [that was very criticised] to the citizen [that was very criticised]*