

LINEARIZATION PREFERENCES GIVEN “FREE WORD ORDER”; SUBJECT PREFERENCES GIVEN ERGATIVITY: A LOOK AT BASQUE

Itziar Laka & Kepa Erdocia
Dept. of Linguistics and Basque Studies
University of the Basque Country, UPV/EHU

ABSTRACT

In this paper, we discuss some recent results from studies of word order (linearization) processing in Basque that show that despite the apparently unconstrained freedom the language displays in linearizing major phrasal constituents in the sentence, native speakers' processing strategies reveal a clear advantage favouring the linearization that directly corresponds to minimal syntactic computation. This is reflected in the time speaker's employ to read sentences, and also in the electrophysiological signals of neural processing measured by means of event related potentials (ERPs). Moreover, when confronted with chains that are syntactically ambiguous regarding two different linearizations, speakers process them as if they were unambiguous, choosing the minimal syntactic computation for their grammar, which is the head final, SOV. We also suggest that the claim that processing preferences universally favour subjects might need to be revisited under the light of ergativity. We will suggest that accounts of processing mechanisms and strategies based on notions like “subject of” or “object of” as explanatory primitives are likely to fail short, and argue that abstracting away from these descriptive categories, and moving into accounts that take into consideration the impact of morphological variability in processing might yield a more accurate view of the interplay of variant and invariant processing mechanisms of language.

1. Introduction: merging theoretical linguistics with experimental methods.

Current linguistics and models of language processing are particularly concerned with understanding the interplay between universal and language specific aspects of linguistic form and language processing strategies. Evidence suggests that syntactic structure is one of the factor that modulates complexity in language processing, though not the only one: variable morphological properties of languages, conceptual features like animacy and lexical frequencies also affect the way in which we process linguistic input.

During the last two decades, there has been a considerable advancement in the discovery of the fundamental design and complex nature of human language, and in the understanding of how language is represented and processed in the brain. There is an increasing trend to combine the fine-grained knowledge of linguistic form and structure provided by Theoretical Linguistics, with the strict and heavily constrained experimental methods of cognitive neuropsychology, in order to reduce the undeterminacy of explanatory models built on data coming solely from external linguistic behavior (Marantz, 2005). More and more, we can inquire into the inner workings of the brain when it performs a linguistic task, and more and more the

questions we can tackle specific issues of computational-combinatorial aspects of language that have been less amenable to experimental work in the recent past.

Thus, Theoretical Linguistics and experimental methods traditionally used in Cognitive Neuropsychology can now be combined to gather more evidence about the neural grammar, and to determine the fundamental underlying properties of language representation processing, by studying a broad sample of languages and phenomena. Cross-linguistic studies of this interdisciplinary type reveal that morphological features can modulate how a language is processed, that is, that linguistic traits and forms have sometimes an impact on the strategies employed in linguistic processing (Bates, Devescovi, & D'Amico, 1999, Bates, Devescovi, & Wulfeck, 2001). For this reason, neurocognitive models of language are becoming increasingly sensitive to the specific morphosyntactic properties of grammars, in order to determine to what degree does linguistic variability have and impact in the way in which language is represented and processed in the human brain (Bornkessel, & Schlesewsky, 2006; Bornkessel-Schlesewsky, & Schlesewsky 2009).

This paper presents and discusses some results recently obtained by combining findings from Theoretical Linguistics with experimental methodologies such as reading times and electrophysiological measurements of brain activity by means of Event Related Potentials (ERPs). It seeks to contribute to unravel how different languages are processed in the human brain, because as we know more about similarities and differences in cross-linguistic processing, we will be able to tell apart common neural underpinnings from other phenomena that might be dependent on language specifications.

2. Is there order when processing free word order?

In Linguistics, the issue of whether free word order languages are fundamentally different in their syntactic structure from fixed word order languages has been thoroughly discussed and intensively studied. Research in the last decades reveals that, despite superficial appearances, when free word order languages are studied in depth evidence for a hierarchical sentence structure emerges, and with it a basic word order that correlates with the linear arrangement of syntactic constituents that results from the structure involving the least syntactic operations. Linguistic research has consistently shown that even grammars with apparently limitless degrees of freedom in surface constituent order have an underlying, canonical word order (Chomsky, 1981; Greenberg, 1963), which typically surfaces in a declarative sentence that initiates discourse, that is, a sentence where no constituent is focalized and where the entire event constitutes new information. Basque is a genuine example of a language with a great degree of freedom of word order. As shown in (1), nearly all constituent permutations are possible (Laka, 1996):

- (1) a. emakume-a-k gaur gizon-a ikusi du
woman-D-erg today man-D seen has
“the woman has seen the man today”
- b. gizona ikusi du gaur emakumeak
- c. gizona ikusi du emakumeak gaur
- d. gaur ikusi du emakumeak gizona

- e. gaur ikusi du gizona emakumeak
- f. emakumeak ikusi du gizona gaur
- g. emakumeak ikusi du gaur gizona
- h. gizona emakumeak ikusi du gaur
- i. gizona gaur ikusi du emakumeak
- j. gaur gizona emakumeak ikusi du
- q. ikusi du emakumeak gizona gaur

Studies in generative grammar in the early eighties investigated and discussed whether languages like Basque, Japanese, Warlpiri and many others that displayed a great variability in surface word order might be *non-configurational*, that is, whether they had a “flat” sentence structure with no hierarchically defined relations among constituents, thus allowing for any order permutation of constituents, or rather, whether despite the surface variation, sentence structure was hierarchically arranged as it is in configurational languages. The hypothesis that some languages were non-configurational was due to Hale (1980, 1983), and it was originally expressed in reference to Japanese and the Australian language Warlpiri, that presented three characteristics that were thought to cluster in *non-configurationality*: free word order, null pronouns and discontinuous constituents. Since the original discussions and research into the possibility that there are configurational and non-configurational languages, most work within the *Principle and Parameters* model in Generative Linguistics has converged in the conclusion that human languages are configurational, that is, that sentence structure is hierarchically arranged (Marác & Muysken 1989). However, work in other approaches such as Lexical Functional Grammar (Austin & Bresnan 1996) argue for the view that non-configurationality a possible trait of human language. Two central issues involved in the configurationality debate are: (a) whether human languages necessarily employ combinatorial, hierarchical structure and (b) whether core grammatical functions like subject and object are dependent on a structural configuration and thus derivative form structure, or whether they are primitives of language, independents from structure.

At the start of the configurationality debate, some authors argued that Basque was a prototypical instance of a non-configurational language (Rebuschi, 1989) but progressively, as more linguists began to study Basque syntax in detail, and more phenomena were considered, linguistic evidence accumulated in favor of hierarchically governed phenomena that were not accountable by the non-configurational hypothesis (Ortiz de Urbina, 1989). Since then, studies on Basque syntax generally agree that SOV is the canonical, basic sentence constituent order of the language, while all other sentential word orders (OSV, SVO and OVS) result from linearizing structures that are derived by means of further syntactic operations (De Rijk 1969, Ortiz de Urbina 1989, Laka 1994, Artiagoitia 1994, Fernandez 1998, Elordieta 2001 among others).

Language processing studies reveal that word-order variations that correlate with syntactic complexity can be detected both behaviorally and electrophysiologically. Canonical, underived sentence word order tends to be processed faster and with greater ease (see Sekerina 2003 for an overview). However, in the case of studies involving free word order languages, results do not always converge in finding that one linearization of constituents is easier and faster to process, and employs less neural resources. For example, in Japanese, studies measuring the time subjects need

to read a set of experimental sentences do not converge in their findings: some report no differences between SOV and OSV word orders (Tamaoka, Sakai, Kawahara, & Miyaoka, 2003; Yamashita, 1997), but others find that OSV imposes higher processing demands (Mazuka, Itoh, & Kondo, 2002; Miyamoto & Takahashi, 2002). More recently, in a study that measures the electrophysiological signals generated by neural processing Hagiwara et al. (2007) did not detect signals to indicate that OSV was costlier to process than SOV. This electrophysiological measurement of brain activity (Event Related Potentials or ERPs) has been rather extensively used to investigate word order processing in German (Bornkessel, Schlesewsky, & Friederici, 2002; Matzke et al., 2002; Rösler, Pechmann, Streb, Röder, & Hennighausen, 1998; Schlesewsky et al., 2003). The studies on German find that object initial sentences generate greater processing efforts, and assume that they arise due to the displacement of the object-phrase, which must wait until its canonical position is reached for full interpretation. This result agrees with previous studies in which differences in word order using relative or interrogative clauses were explored in this language (Felser et al., 2003; Fiebach et al., 2002; King & Kutas, 1995; Kluender & Kutas, 1993; Müller et al., 1997; Münte, Heinze, Matzke, Wieringa, & Johannes, 1998). Thus, it appears that, at least for German, non-canonical sentences require a larger integration cost and make greater demands on working memory. These results also converge with linguistic accounts of German syntax, where object-initial main sentences involve displacing the object from its canonical position to a higher place in the sentence-structure (Schwartz & Vikner, 1996). To summarize, evidence from Japanese is not conclusive regarding a processing advantage for canonical versus non-canonical word order processing in declarative sentences, but evidence from German reflects a strong preference for canonical orders, showing that object-initial sentences are costlier to process.

If (a) syntactic complexity can be detected by means of reading times and electrophysiological measures, and if (b) the SOV linearization corresponds to the syntactically less complex sentence in Basque, then (all other things equal) a processing asymmetry should obtain favoring SOV in this language as compared to all other word orders in declarative sentences. This processing advantage should be detected in the form of shorter reading times for SOV, and electrophysiological indexes of processing complexity should not arise in SOV as compared to other word orders.

In Erdocia et al. (2009) we investigated this question, to determine whether these effects obtained in the case of verb final sentences. The study compared the processing of SOV versus OSV sentences like (2); it also explored a particular case of full syntactic ambiguity that has not been explored before for any language such as the examples below in (7). A set of experimental sentences was created that could be parsed either as SOV or as OSV, and thus they provided the opportunity to determine what the parser chose to do in the absence of any other information or cue. That is, it allowed to determine whether subjects showed a marked preference for one of the two possibilities in the absence of any other cue that indicated how to parse the sequence; in other words, it provided a unique window to see what the parser does morphology is blind and linearized syntactic structure is the only cue available. The results showed that SOV sentences had a clear processing advantage in reading times, in electrophysiological measurements, and in ambiguity resolution, thus revealing that linguistic accounts of Basque syntax regarding word order and syntactic complexity reflected facts about

neural computation, and also that the minimal hierarchical structure of sentences in this language is detectable by both behavioral and ERP methods.

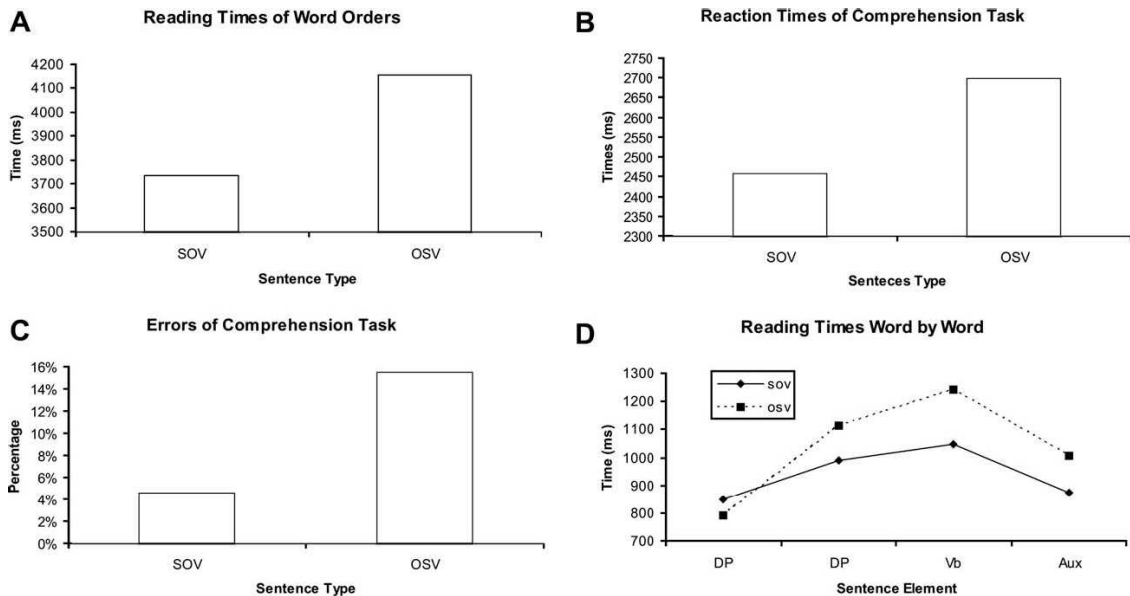
The first experiment we undertook measured how long it took to read a set of experimental sentences to a group of 23 native speakers. The sentences were presented word by word (each of the words presented was also a phrase, since Basque is an agglutinative language) in a computer screen. Participants pressed the bar once they had read the phrase and the next phrase appeared until completing the sentence¹. The lexical elements used in the experimental sentences were controlled for length and frequency. Each participant was presented with 16 SOV and 16 OSV sentences that were never repeated, and mixed with the experimental sentences they also read 32 unrelated sentences that were matched in the number of words/phrases in order to distract them from the experimental condition. The experimental sentences were like those shown in (2):

- (2) a. emakume-a-k gizon-a ikusi du SOV condition
 woman-D-erg man-D seen has
 ‘the woman has seen the man ’
- b. gizon-a emakume-a-k ikusi du OSV condition
 man-D woman-D-erg seen has
 ‘the woman has seen the man ’

After reading each sentence, the participant read a yes/no comprehension question on the screen, to assure attention to the experimental task, and also to provide indication of sentence-comprehension. For instance, for the sentences in (2), the question would be “is it true that a woman has seen a man?”. We took these responses into consideration, and also measured how much time it took them to provide an answer. The results from this first experiment are shown in (3):

¹ Details of the participants, materials and procedure in Erdocia et al. (2009). Here we only summarize the most general aspects of the experimental conditions and focus on the discussion of the results.

(3) Results from Experiment 1: reading and understanding SOV versus OSV (from Erdocia et al., 2009)



As shown in (3A), the total time participants required to read the sentences was significantly longer for OSV sentences and shorter for SOV. (3B) shows that participants took longer to read and answer comprehension questions related to OSV as compared to SOV. (3C) shows that, even though the amount of errors in the comprehension question was very low, it was significantly higher for OSV sentences. All these measures indicate that processing an SOV sentence is significantly faster and easier than processing OSV sentences for native speakers of Basque, even though all sentences are equally grammatical. This result is convergent with the claim made in Linguistics that OSV sentences are syntactically more complex than SOV sentences in this language (De Rijk 2007; Hualde & Ortiz de Urbina 2003). Thus, the results from our first experiment are consistent with the predominant view in Theoretical Linguistics regarding word order in Basque, which states SOV as canonical, and all others, including OSV, as structurally more complex (De Rijk, 1969; Ortiz de Urbina, 1989).

3. How are subjects processed in an ergative language?

The diagram in (3D) shows the reading times required for each phrase in the sentence in each condition. As can be seen, phrases in the canonical condition took significantly shorter than phrases in the derived condition. Except in one case: participants always took longer reading the subject DP than they took reading the object DP. This result is novel in processing studies and diverges sharply from what is usually obtained with the same method (self-paced reading) in other languages such as Dutch (Kaan, 1997) and Russian (Sekerina, 1997) where subjects are always processed faster than objects. Hence, the result merits some discussion, because it is due to a characteristic morphological property of Basque grammar: *ergativity* (Dixon 1994). Unlike all languages previously studied in processing research wither behaviourally or electrophysiologically, Basque is not a nominative-accusative language, and it is customarily described as falling in the *ergative* class (Levin 1983, de Rijk 2007).

In ergative grammars, intransitive subjects and objects belong in one morphological class (known as *absolutive*) which is unmarked, and transitive subjects are marked and belong in a separate morphological class (known as *ergative*). This is shown in (4), where the transitive subject *emakumeak* “the woman” in (4a) is marked with ergative case (*-k*), whereas the object *gizona* “the man” carries no overt case marker, as the intransitive subject in (4b).

(4) a. emakume-a-k gizon-a ikusi du
woman-D-erg man-D seen has
“the woman has seen the man”

b. gizon-a etorri da
man-D arrived is
“the man has arrived”

Thus, unlike nominative languages where subjects tend to be unmarked, in ergative grammars the transitive subject carries more overt morphology than the object. This is probably one of the factors at play in the longer reading times for subject DPs obtained in this experiment. Another factor that must be kept in mind, related to ergativity, is that, when participants encountered the object DP in the experimental sentence (2b) they could interpret it either as an intransitive/unaccusative subject or as a sentence-initial object. If Basque speakers employ a “subject-first” processing strategy (Frazier and Fodor 1978; Bates et al. 1988, Bornkessel & Schlesewsky 2006), a strategy that has been widely reported in all previous processing studies, then they should interpret that initial object DP as an intransitive subject, which could also contribute to explaining the faster reading times.

Another factor that might play a role in the peculiar reading times obtained for subject and object DPs in this sentence processing experiment does not have to do with a typological trait, but with a specific morphological ambiguity found in Basque. This ambiguity involves the *-ak* ending on DPs, which we illustrate in (5). The ending is homophonous to a singular transitive subject morphology and a plural object morphology. In each of the cases, the morphological structure of the DP is different, though the words look alike. Plurality in Basque is marked solely in the Determiner, which can be singular *-a*, as shown in (5a), or plural *-ak*, as shown in (5b)². Determiner Phrases must be marked with ergative case when they are transitive subjects, and the form of the ergative case marker is *-k*. As shown in (5c), the result of adding ergative case to a singular DP yields the sequence *-ak*, which is homophonous to the plural determiner in (5b). For completeness, (5d) shows the resulting form of merging the plural determiner *-ak* with the ergative marker *-k*, which is the ending *-ek*, unambiguously denoting a plural ergative NP.

² We leave aside other elements pertaining to the Determiner category, such as demonstratives, which are not relevant to this discussion. The Determiner *-a*, *-ak* is not always translationally equivalent to English *the*. For further details on the determiner system and the structure of the DP in Basque see Artiagoitia (2006).

- (5) a. emakume-a
woman-D
“the woman”
- b. emakume-ak
woman-D_{pl}
“the women”
- c. emakume-a-k
woman-D-erg
“the woman (ergative case)”
- d. emakume-ek
woman-D_{pl}+erg
“the women (ergative case)”

Hence, upon encountering as input something like *emakumeak*, two possible interpretations are compatible with Basque grammar: interpreting it as a singular ergative meaning “the woman”, or interpreting it as a plural absolutive meaning “the women”. If we now turn to the examples of experimental sentences shown in (2), we see that when participants encountered the sentence initial subject DP in (2a), they could in principle process it in three different ways:

- (i) as a transitive singular subject,
- (ii) as an intransitive plural subject,
- (iii) as a sentence initial plural object.

Similarly, and given the head-final setting of Basque, when they encountered the subject DP as the second phrase in (2b), they could interpret it as:

- (i) a transitive singular subject of the main clause,
- (ii) an intransitive plural subject of an embedded sentence,
- (iii) a plural object of an embedded clause.

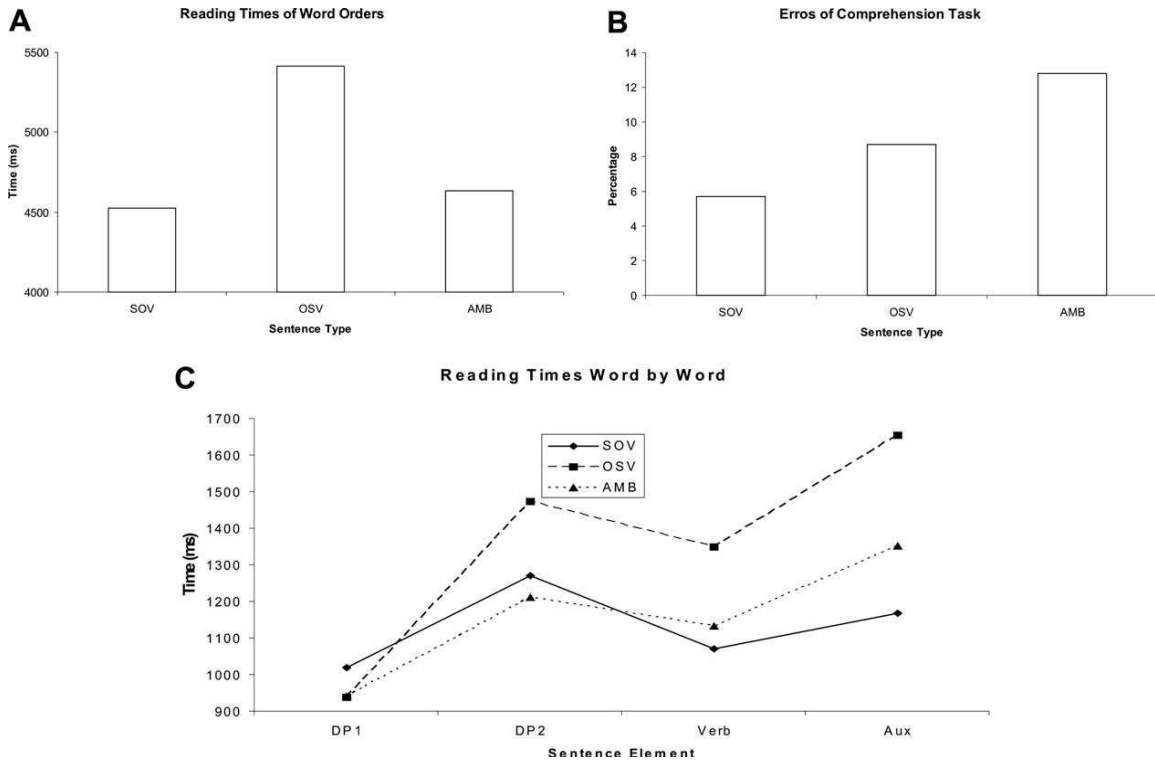
At this point, we cannot yet determine which one of these factors, and to what degree, are behind the longer reading times obtained for subjects in transitive sentences in Basque. But as we will see, this divergent and novel pattern in subject/object processing costs will remain constant across different experimental conditions.

4. Further investigating linearization and subject preference.

Given these initial results, we designed a second experiment in order to (a) see whether the advantage for SOV sentences would also be maintained for sentences where both DPs were plural, (b) explore what the processing preferences of participants would be when confronted with sentences that could be equally parsed as either SOV or as OSV. In order to do this, we created transitive sentences like those shown in (6):

- (6) a. emakume-ek gizon-ak ikusi dituzte
women-D_{pl}/erg man-D_{pl} seen have
“The women have seen the men”
- b. gizon-ak emakume-ek ikusi dituzte
man-D_{pl} women-D_{pl}/erg seen have
“The women have seen the men”

(8) Results from Experiment 2: reading and understanding plural SOV, plural OSV and Ambiguous sentences (from Erdocia et al 2009).



Let us consider the reading times for the sentences, shown in (8A): participants took the same time to read the SOV sentences as they took to read the fully ambiguous ones. This indicates that they were not treating the ambiguous sentences any different from the unambiguously SOV ones, that is, the bias towards processing fully ambiguous sentences resorting to the simplest syntactic computation, namely SOV is so strong that they were not even aware of the ambiguity. In contrast to this, the OSV sentences took considerably longer to read, indicating an increase in processing difficulty.

5. What does the subject-preference strategy prefer in an ergative language?

Consider now the times employed to read each constituent in experiment 2, and in particular let us consider the relative reading times of the initial DPs of the sentences (see 8C). The SOV condition, depicted with the solid line, is the one where participants initially encountered an unambiguously plural ergative/subject DP like *emakumeek* “the women”. Interestingly, despite it being unambiguously a subject, it takes longer to read than the other two cases, where the first DP encountered had the form *emakumeak*, which is ambiguous between a singular ergative (subject) “the woman” or a plural absolutive (intransitive subject or object) “the women”. Even in the face of this ambiguity, participants are faster processing this DP than they are processing the plural transitive subject. Given the strictly incremental nature of language processing, this indicates that speakers are not initially aware of the ambiguity and make a rapid choice between the two possibilities.

On the other hand, the phrase-by-phrase reading times for the SOV sentences and the ambiguous ones is practically identical, strongly suggesting that participants are not aware of the ambiguity while reading the sentences, a result that converges with the total reading times of SOV and ambiguous sentences. That is, both types of sentences, ambiguous and unambiguous, are processed as if they were the same.

This in turn means that the initial DP in the ambiguous condition is interpreted as a *singular ergative* out of the two possible choices, and therefore, it also means that speakers are not entertaining the possibility that the ambiguous DP is a plural absolutive. However, an absolutive form is a viable subject in an ergative language like Basque. In fact, the simplest, syntactically most economical subject in an ergative language is the absolutive, for this is the morphological form associated to unaccusative predicates, that is, to truly monoargumental predicates. It therefore appears that subject preference in Basque is biased towards an *agent* preference, which is the ergative (Laka 2006) rather than being biased to the syntactically most economical subject, which is the absolutive/theme subject of unaccusatives (Levin 1983).

6. Electrophysiological correlates of linearization and subject preferences.

In a third experiment, Erdocia et al. (2009) sought to obtain electrophysiological evidence of the effects observed in the self paced reading experiments 1 and 2. We studied data from 24 native speakers of Basque that took part in the experiment. The experimental materials consisted of two groups of sentences. In the first group, there were sentences like (9):

- (9) a. otso-ek ardi-ak jan dituzte b. ardi-a otso-a-k jan du
 wolf-D_{erg.pl} sheep-D_{pl} eaten have sheep-D-erg wolf-D-erg eaten has
 “The wolves have eaten the sheep” “The wolf has eaten the sheep”

This experimental condition was aimed at determining whether the comparison between the initial DPs of these sentences, generated any electrophysiological response of the type that has been previously observed in German by Matzke et al. (2002), and Rösler et al. (1998) or not, as reported for Japanese by Hagiwara et al. (2007). That is, we wanted to find out whether comparing the electrophysiological correlates elicited at the processing time of those sentence-initial arguments, the Object of non-canonical OSV sentences, versus the Subject of canonical SOV sentences (examples 9b and 9a respectively) revealed signs of processing complexity for the initial DP in (9b), of the type that has been reported in the literature for syntactically displaced constituents.

The second group of experimental sentences consisted of syntactically ambiguous sentences where the DPs were marked with *-ak*, formally similar to (7), but in this occasion, both readings of the sentences were not compatible with the speaker’s world knowledge. Examples of these sentences are provided in (10):

- (10) a. otso-a-k ardi-ak jan ditu c. otso-ak ardi-a-k jan ditu
 wolf-D-erg sheep-D_{pl} eaten has wolf- D_{pl} sheep-D-erg eaten has
 “the wolf has eaten the sheep” “the sheep has eaten the wolves”
- b. ardi-ak otso-a-k jan ditu d. ardi-a-k otso-ak jan ditu
 sheep-D_{pl} wolf-D-erg eaten has sheep-D-erg wolf- D_{pl} eaten has
 “the wolf has eaten the sheep” “the sheep has eaten the wolves”

The externalized forms corresponding to sentences (10a,b) are the same, and the externalized forms corresponding to sentences (10c,d) are also the same. In this sense, the syntactic ambiguity of these examples is exactly the same as the one illustrated in (7). However, unlike the set of sentences employed in experiment 2, where the both grammatically possible interpretations were equally plausible, the sentences in this third experiment had a different quality: one of the interpretations was plausible, like for instance (10a) and (10b), but the other one was not, because it goes against our world knowledge, like (10c) and (10d). Participants were presented with an equal number of sentences that were plausible in the SOV interpretation but not in the OSV interpretation, and vice-versa, sentences that were plausible in the OSV interpretation but not in the SOV interpretation.

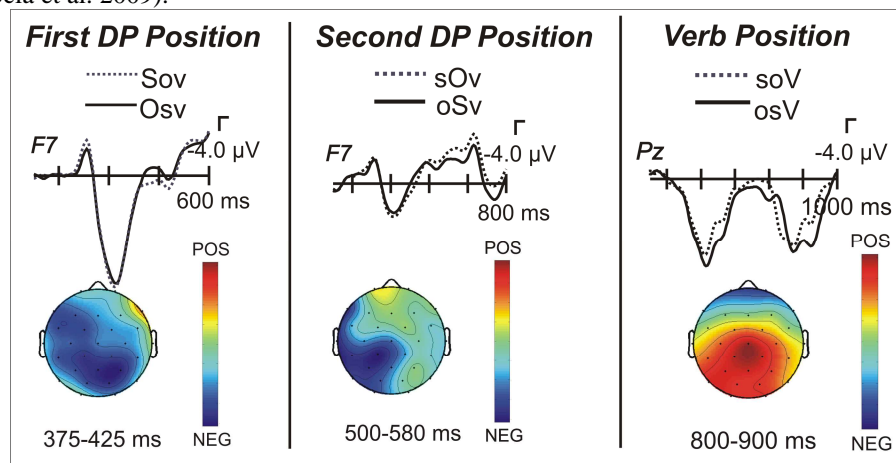
Thus, using sentences like (9a,b) and (10a,b) we generated 240 different sentences in order to conduct an ERP experiment. Twenty-four native speakers of Basque gave informed consent to participate in the ERP experiment. None of the participants had previous neurological history and all had normal or corrected-to-normal vision. The participants were all right handed according to the Edinburg Handedness inventory (Oldfield, 1971). All participants read 60 sentences of each experimental condition (9a,b and 10a,b). Sentences were presented word by word in the middle of a computer screen. Each word lasted 250 milliseconds on the screen and the interval between words was 250 milliseconds. Participants had to keep their eyes open until the end of each sentences, then they had 3000 milliseconds for eye blinking. After that, following a fixation point which remained for 1500 milliseconds, next sentence began. Every eight sentences participants had to provide an answer to a simple question related to one of the previous eight sentences. This task was used in order ensure participants were paying attention to the sentences. Each ERP recording session lasted about 45–50 min; participants were told that the main purpose of the experiment was to read carefully the sentences presented and to answer correctly the questions related to the sentences.

ERPs were recorded from 29 positions in the scalp using tin electrodes mounted in an electrocap. EEG data was rereferenced off-line to the mean of the activity at the two mastoid processes. Vertical and horizontal eye movements were monitored with an electrode at the infraorbital ridge and another electrode at the outer canthus of the right eye. Electrode impedances were kept below 5 kΩ. The electrophysiological signals were filtered and digitized. Trials with base-to-peak electro-oculogram (EOG) amplitude of more than 50 μV, amplifier saturation, or a baseline shift exceeding 200 μV/s were automatically rejected off-line. Percentage of artifact rejection was 6.7%. Stimulus-locked ERPs were averaged for epochs of 1024 ms starting 100 ms prior to the stimulus. First, two omnibus repeated measures ANOVAs were conducted for the initial evaluation of the stimulus-locked ERP activity in two locations. In order to decompose the interactions encountered in the omnibus analysis and to have a finer-grained

analysis, further pairwise ANOVAs were conducted. For all statistical effects involving two or more degrees of freedom in the numerator, the Hynh–Feldt epsilon was used to correct for possible violations of the sphericity assumption (for more details about EEG recording and data analysis, see Erdocia et al., 2009).

Responses to the questions asked showed that participants performed the experiment very well. The comprehension task showed a mean percentage of correct responses of 91%. The main ERP results confirmed the evidences obtained in the previous behavioral experiments with regard to the higher processing cost derived from the higher syntactic complexity of non-canonical Object-first sentences (Experiment 1), and with regard to the existence of a default preference for an SOV linear order of constituents in the sentence observed in Basque, which is also applied in the case of fully ambiguous sentences (Experiment 2). Regarding the initial position in the unambiguous sentences (SOV vs. OSV), the ERPs showed an increased negativity at object position. Similar negative result were found at sentence second position for the subject of OSV when comparing with the object of SOV. Finally, when comparing unambiguous sentences at verb position, a clear P600 component was observed:

(11) Results from Experiment 3: ERP signals of the unambiguous SOV and OSV sentences (adapted from Erdocia et al. 2009).



As in the case of object-first sentences in German (Felser et al., 2003; Matzke et al., 2002; Rösler et al., 1998), we can conclude that syntactically fronted/displaced constituents generate electrophysiological differences because the structure of the sentences with displaced constituents is more complex and requires further processing operations.

Regarding the processing of syntactically ambiguous sentences (10a,b) no significant differences were found when comparing subjects and objects at first and second sentence positions, suggesting that the sentence initial ambiguous DP was interpreted as the ergative singular subject of the sentence, and the second ambiguous DP was interpreted as the absolutive plural object. These results converge with the results from experiment 2 in which we used fully ambiguous sentences. Interestingly, at verb position of sentences in which we forced a reinterpretation of first two arguments from Subject-Object to Object-Subject alignment by means of the world knowledge of participants (10b), we observed a long lasting negativity reflecting the working memory load associated to syntactic reanalysis.

Our electrophysiological results provide evidence from neuroimaging measures supporting that even languages displaying free word order have an underlying canonical order, corresponding to the least complex sentence-structure.

7. Subject preference vanishes in verb medial, non-canonical sentences

Regarding verb final sentences, canonical order is processed faster and with greater ease even in a highly inflected and free word order language like Basque; we now focus on verb medial SVO and OVS sentences in order to see whether there is a default processing mechanism favouring subject-first sequences (Erdocia et al. in press). For that purpose, we designed a behavioral experiment comparing verb medial subject-first sentences like (12a,c) with object-first sentences like (12b,d).

- | | | |
|------|---|--|
| (12) | a. emakume-a-k ikusi du gizon-a
woman-D-erg. seen has man-D
'the woman has seen the man' | b. emakume-ek ikusi dituzte gizon-ak
woman-D _{erg.pl.} seen have man-D _{pl.}
'the women have seen the men' |
| | c. gizon-a ikusi du emakume-a-k
man-D seen has woman-D-erg.
'the woman has seen the man ' | d. gizon-ak ikusi dituzte emakume-ek
man-D _{pl.} seen has woman-D _{pl.erg.}
'the women have seen the men' |

Further, we also included fully ambiguous sentences like (13) which could be interpreted either as SVO or OVS sentences:

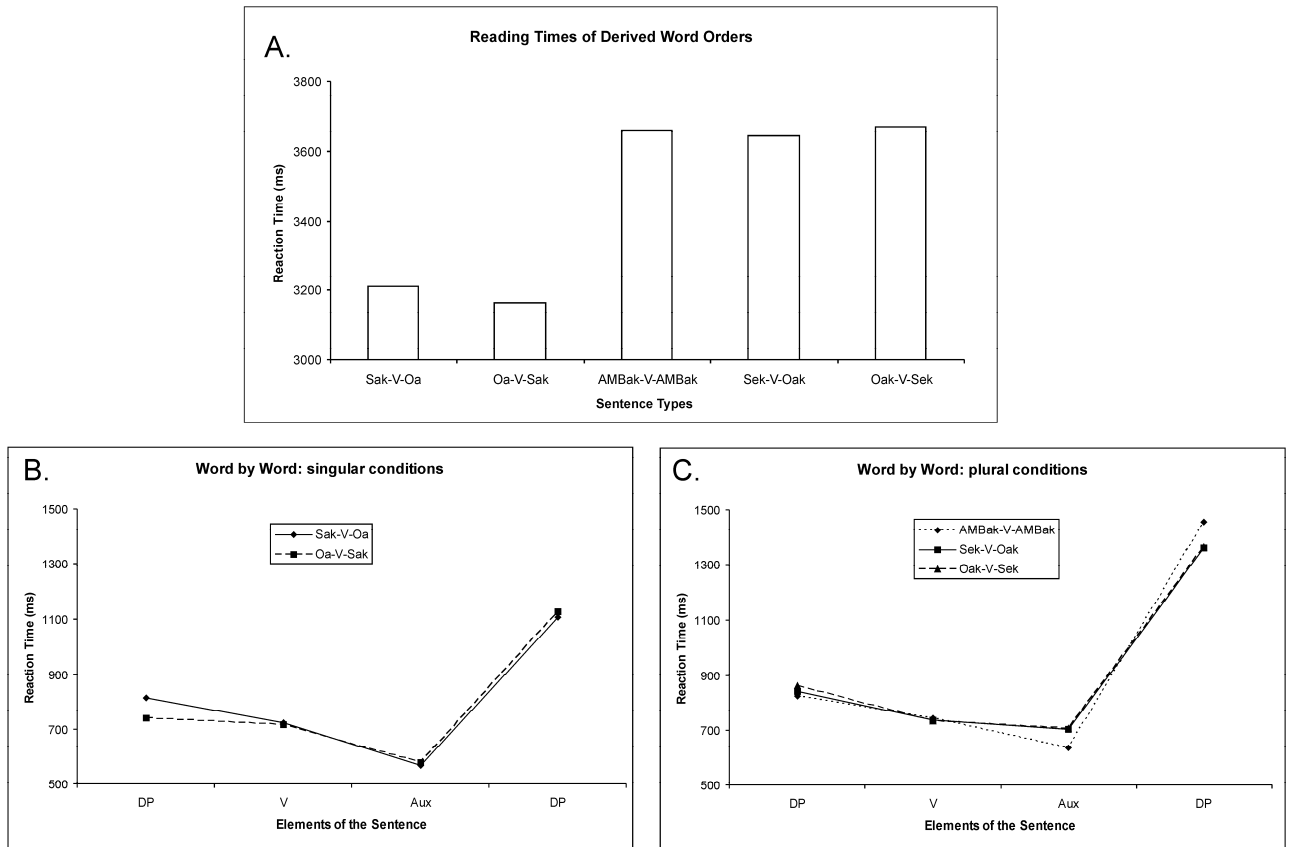
- | | | |
|------|--|--|
| (13) | a. emakume-a-k ikusi ditu gizon-ak
woman-D-erg. seen has man-D _{pl.}
'the woman has seen the man' | b. emakume-ak ikusi ditu gizon-a-k
woman-D _{pl.} seen has man-D-erg.
'the man has seen the woman' |
|------|--|--|

Recall that DP constituents carrying the morpheme *-ak* are ambiguous. Therefore, sentences (12a,d) and (13) are ambiguous at first DP constituent position. Sentences (12a) and (12d) are disambiguated when the inflected auxiliary is reached. In the case of (12a) the sentence unfolds as SVO and in the case of (12d) the resulting structure is OVS. Only sentences like (13) are fully ambiguous because both constituents carry the *-ak* ending (see also examples in 5).

The experimental materials consisted of 100 different sentences like (12a). Starting from this type of sentence, we generated the rest of conditions by means of changing the order and the number of the constituents. The experimental material was randomly mixed with 100 of filler sentences. For this experiment we used the same method used in experiments 1 and 2.

The results from twenty four native speakers of Basque who participated in the experiment are shown in (14):

(14) Results from Experiment 4: Reading and understanding verb-medial sentences in Basque (from Erdocia et al., in press).



Comparing the plural and the ambiguous conditions, a significant difference was observed at auxiliary position in (14C). That difference could be due to fact that ambiguous auxiliary was shorter in length or less complex morphologically (*d-it-u* “3rd person/object-plural-root”) than plural auxiliary (*d-it-u-zte* “3rd person/object-plural-root-plural/3rd person/subject”). Whatever the factors responsible for the differences between auxiliaries, syntactic structure is not one of them.

The results revealed that verb-medial SVO and OVS sentences do not display a processing asymmetry. As it is shown in (14A), the only significant differences obtained in the comparison between singular and plural conditions. Those differences were arguably due to conceptual and/or morphological complexity, but clearly not to syntactic structure. Considering the word by word reading times of singular conditions shown in (14B), we replicated the previous finding of experiment 1 where absolute constituents were read faster than ergative constituents. However, the rest of the sentence was processed equally in both SVO and OVS conditions.

These results converge with claims made in Basque Linguistics that all word orders other than SOV are derived via syntactic displacement. However, studies of word order processing in German, another SOV language, report an advantage of Subject initial (SVO) versus Object initial (OVS) orders. In German, where canonical order is SOV, and V2 applies in main sentences, Bader and Meng (1999) showed that participants entered a garden-path in OSV temporally ambiguous structures and favored the SOV

interpretation. Using ERPs, Rösler et al. (1998) showed that sentences whose word order deviates from the SOV canonical structure were more difficult to process (see also Bornkessel et al., 2002, Schlesewsky et al., 2003 among others). In verb medial sentences, ERPs differentiated between SVO and OVS structures in German (Matzke et al., 2002). In Basque, behavioral and ERP measurements differ significantly for canonical SOV and derived OSV word orders (Erdocia et al., 2009): reading times showed longer reading times and syntactic reanalysis effects for OSV structures (see also Fig. 10.2), and ERPs showed negativities in both DPs and P600 at verb position suggesting that derived OSV word order required further computational resources (Erdocia et al., 2009; Fig. 10.3 in this chapter). Thus, while German and Basque reveal a similar processing advantage for SOV versus OSV, an advantage that has also been reported for Japanese, also SOV³, processing of verb-medial word orders yields diverging results: in German, a Subject-initial advantage has been reported, whereas in Basque both Subject-initial and Object-initial orders reveal equivalent measures of processing complexity. Arguably, this difference might be due to the fact that German is a V2 language, whereas Basque is not. In V2 languages, main sentences have the inflected auxiliary or verb always in second position in the sentence, and the most frequent constituent preceding the inflected verb is the Subject; this entails that SVO is in fact a rather frequent word order for a V2, main sentence in German, whereas this is not the case for Basque. This explanation is consistent with the results from German, showing a preference of SOV versus OSV, and then a preference for SVO versus OVS, whereas Basque displays a preference for SOV versus OSV, but no preference in the case of SVO and OVS.

We did not find any signs of a processing advantage for the Subject initial order (SVO) as compared to the Object initial order (OVS). The experiment 4 suggests that both derived word orders required similar computational resources, with no advantage for the subject-first sequences. On the other hand, our data did not indicate any ambiguity resolution process reading temporally ambiguous sentences and fully ambiguous chains.

8. Concluding discussion.

In this paper we have discussed the consequences of some recent findings from ERP studies on sentence processing in Basque. One central issue we have addressed is the relationship between syntactic complexity and processing complexity, in particular whether there is evidence for a processing mechanism that favours a choice for the word order corresponding to the least complex syntactic structure in a heavily inflected and free word order language. Another central issue is to determine whether there is evidence for the so-called “subject-preference” processing strategy in an ergative language like Basque, where the class of subjects is not identified in morphology like it is in nominative languages.

It is well established by now that some aspects of processing complexity do not derive from syntactic complexity, such as the contrast between nested and non nested structures, for instance. However, syntactic complexity has a significant impact in processing, and hence diverging processing effects obtain depending on the syntactic

³ In Japanese whose structure is always verb-final some found differences between SOV and OSV structures (Miyamoto & Takahashi, 2002; Mazuka et al., 2002; Wolff et al., 2008), but some others did not find any difference (Yamashita 1997; Tamaoka et al., 2003; Hagiwara et al., 2007).

representation processed. This was found to be true also in Basque, as shown by the increasing processing costs ensued by various linear orders of syntactic constituents in sentences that departed from canonical SOV order. That is, favoring the processing route towards minimal syntactic structure, we find that canonical, basic word order is processed faster and with greater ease, presumably because it involves the simplest neural computation. Therefore, even in highly inflected, free-word order languages like Basque, where other processing cues could be thought to take precedence over syntactic structure, experimental evidence indicates that (a) the canonical constituent linear order of the sentence that results from minimal syntactic computation is the easiest and preferred choice for the parser, and (b) processing relies strongly on syntactic structure in sentence comprehension even when other cues such as case and verbal inflection are available. Given similar findings in other languages (English, German, Japanese, etc) this appears to be a universal design property of language, regardless of morphological differences and other variable specifications of linguistic form in a given grammar.

Regarding the subject preference strategy for processing, these results we have discussed are novel and diverge sharply from what is usually obtained with the same method (self-paced reading) in other languages such as Dutch (Kaan, 1997) and Russian (Sekerina, 1997), where Subjects are always processed faster than Objects, and in the face of ambiguity subject preference prefers intransitive subjects (Cf. Turkish Bornkessel-Schlesewsky 2007). In the word order processing studies reported here for Basque, we observe that Subject DPs always take longer than Object DPs to process (and it is subjects rather than objects that generate frontal negativities in ERPs), a result that has not been reported in the literature so far, where nominative/accusative languages were studied exclusively.

Let us further consider the findings in the light of the subject-preference hypothesis. Given an ambiguous DP-*ak* of the kind shown in examples (5b,c), which can be interpreted either as a singular ergative DP or as a plural absolutive DP, note that both choices of parsing would reflect a “subject-preference” in an ergative language, because both interpretations yield a possible agentive subject in the case of the ergative, and an unaccusative, thematic subject in the case of the absolutive. Put it differently, any available interpretation of an initial ergative or absolutive DP is consistent with a subject preference in an ergative language like Basque. However, our data reveal a strong preference for one of the two choices, that is, a specific *type* of subject-preference, in the sense that speakers prefer to parse the initial DP as an *agentive* subject (transitive or unergative), but not as a thematic subject (unaccusative).

In the processing literature, the subject preference strategy has been found to strongly determine speakers’ processing choices in several languages. The subject preference is revealed for instance in a preference to parse an ambiguous first DP as subject. It is generally assumed that this strategy is one particular example of the more general “minimal effort” nature of the parser (Frazier & Fodor, 1978; Gorrell, 1995; de Vincenzi, 1991). Notions like “minimality” or “economy” play a central role in contemporary linguistics, and have been argued to be a driving force in the architecture of language (Chomsky 1989, 1995; Rizzi, 1990). In the processing literature, the idea that economy of efforts favours minimal constructs is also found early on (Miller 1962; Miller & Chomsky, 1963; Miller & McKean, 1964), and pervades the literature. One recent formulation of this minimal effort nature of the parser is found in Bornkessel & Schlewsky, 2006, who state it as follows:

Minimality: In the absence of explicit information to the contrary, the human language comprehension system assigns minimal structures. This entails that only required dependencies and relations are created.

(Bornkessel & Schlesewsky, 2006:790)

This formulation, like any other whose driving force is to minimise the syntactic complexity expected for a given linguistic input. Regarding the subject-preference processing mechanism, minimalist accounts make the prediction that unaccusative subjects should be favored over unergative or transitive subjects in language processing, all other things being equal. Put differently, minimal effort principles require that in the absence of explicit information to the contrary, the human language comprehension system favor an unaccusative subject reading over a transitive or unergative subject reading.

However, the processing data obtained from Basque seem to run contrary to this prediction. Instead, what we find is that speakers prefer to interpret the ergative/absolute ambiguous DP as ergative, that is, as a transitive/unergative subject, rather than entertaining the possibility that it is an absolute DP, which is the form of unaccusative subjects. In other words, there appears to be an “agent-preference” rather than a “subject-preference”, such that it overrides the choice of an unaccusative subject when both an agentive and a thematic reading of the subject are possible. Whether this is the right interpretation for this processing preference must be confirmed by future research.

Cross-linguistic investigations are crucial to discover the ultimate source of processing asymmetries, and to differentiate between universal processing mechanisms and the impact of grammatical properties of the languages at play. Therefore, in order to understand the nature of complexity in language processing, a broad sample of different grammars must be studied, so that the results of this cross-linguistic research reveal the mechanisms at play in language processing at an adequate level of abstraction. In this vein, the results from these investigations into processing phenomena in Basque show that certain previous findings in ERPs language-processing studies observed in head initial, nominative and relatively fixed word order languages, are also observed in a head final, ergative, free word order and highly inflected and ergative language like Basque, which strongly suggests that these findings signal universal processing mechanisms, independent of parametric specifications of the grammars at stake. However, accounts of processing mechanisms and strategies based on notions like “subject” or “object” as universal primitives of language might not properly account for the fine-grained nature of results from cross-linguistic studies in language processing. Abstracting away from these descriptive categories, which have been argued to be derivative in generative grammar since Chomsky (1965), and moving into more sophisticated and detailed linguistic analysis will plausibly yield a more accurate view of what the invariant, underlying processing mechanisms of language might be.

ACKNOWLEDGEMENTS:

This research has been supported by grants from the Spanish Ministry of Education and Science within the program CONSOLIDER-INGENIO 2010 (BRAINGLOT CSD2007-00012), the Spanish Ministry of Science and Innovation (FFI2009-09695), and the Basque Council for Education, Universities and Research (IT414-10)

REFERENCES:

- Artiagoitia, X. (1994) *Verbal Projections in Basque and Minimal Structure*. Supplements of *Anuario del Seminario de Filología Vasca Julio de Urquijo: International journal of basque linguistics and philology*, Vol. 28, 2, 339-504
- Artiagoitia, X (2006) Basque adjectives and the functional structure of the noun phrase J. Lakarra and J.I. Hualde (eds) *Studies in Basque and historical linguistics in memory of R. L. Trask*. Supplement of *Anuario del Seminario de Filología Vasca Julio de Urquijo: International journal of basque linguistics and philology*, Vol. 40, 1-2, 107-132
- Austin, Peter and Joan Bresnan. (1996) Non-configurationality in Australian aboriginal languages." *Natural Language and Linguistic Theory* 14, 215–268.
- Bates, E., Friederici, A., & Juarez, L. (1988) On the preservation of word order in aphasia. *Brain and Language*, 33(2), 323–364.
- Bates, E., Devescovi, A., & D'Amico (1999) Processing complex sentences: A cross-linguistic study. *Language and Cognitive Processes*, 14(1), 69-123.
- Bates, E., Devescovi, A., & Wulfeck, B. (2001) Psycholinguistics: A cross language perspective. *Annual Review of Psychology*, 52, 369–396
- Bornkessel, I., Schlesewsky, M., & Friederici, A. (2002) Grammar overrides frequency: Evidence from online processing of flexible word order. *Cognition*, 85, B21–B30.
- Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4), 787–821.
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The Role of Prominence Information in the Real Time Comprehension of Transitive Constructions: A Cross-Linguistic Approach. *Language and Linguistics Compass* 3, 19–58.
- Chomsky (1965) *Aspects of the Theory of Syntax*. Cambridge MA: The MIT Press.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Foris: Dordrecht.
- Chomsky, N. (1989): “Some notes on economy of derivation and representation” in I. Laka and A.K. Mahajan (eds.) *Functional Heads and Clause Structure MITWPL* 10, Cambridge MA, 43-74.
- Chomsky (1995). *The Minimalist Program*. Cambridge, MA: The MIT Press.
- De Rijk, R. (1969). Is Baque an S.O.V. language? *Fontes Linguae Vasconum*, 1–3, 319–351 [Reedited by De Rijk, Rudolf P. G., 1998, *De Lingua Vasconum: Selected Writings*. Supplements of the Anuario del Seminario de Filología Vasca “Julio de Urquijo” XLLIII].
- De Rijk, R. (2007). *Standard Basque, a progressive grammar*. Cambridge MA: MIT Press.
- de Vincenzi, M., (1991). *Syntactic Parsing Strategies in Italian*. Kluwer, Dordrecht.
- Elordieta, A. (2001). *Verb Movement and Constituent Permutation in Basque*. Utrecht: LOT Dissertation Series.
- Erdocia K., Laka I. Mestres 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, 1-7

- Erdocia K., Laka I., and Rodríguez-Fornells A. (in press) Processing verb medial word orders in a verb final language. In Peter de Swart and Monique Lamers (eds.). *Case, Word Order, and Prominence. Interacting Cues*, Studies in Theoretical Psycholinguistics series, Dordrecht/Berlin: Springer.
- Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. *Brain and Language*, 87, 345–445.
- Fernández, B. (1998) *Egiturazko kasuaren erkaketa euskaraz* Servicio Editorial de la Universidad del País Vasco, Bilbao.
- Fiebach, C. J., Schlesewsky, M., & Friederici, A. (2002). Separating syntactic integration cost during parsing: The processing of German WH-questions. *Journal of Memory and Language*, 47, 250–272.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6, 291–325.
- Gorrell, P., 1995. *Syntax and Parsing*. Cambridge University Press, Cambridge.
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of language* the order of meaningful elements. Cambridge, MA: MIT Press, pp. 73–113
- Hagiwara, H., Takahiro, S., Masami, I., & Imanaka, K. (2007). A topographical study on the event-related potential correlates of scrambled word order in Japanese complex sentences. *Journal of Cognitive Neuroscience*, 19, 175–193.
- Hale, Kenneth L. (1980). Remarks on Japanese phrase structure: Comments on the Papers on Japanese Syntax. In Y. Otsu & A. Farmer (Eds.), *Theoretical Issues in Japanese Linguistics MITWPL 2*, 185-203.
- Hale, Kenneth L. (1983). Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory*, 1, 5-47.
- Hualde, J. I., & Ortiz de Urbina, J. (Eds.). (2003). *A Grammar of Basque*. Berlin: Mouton de Gruyter.
- Kaan, E. (1997). *Processing subject–object ambiguities in Dutch*. Ph.D. Thesis: University of Groningen (Groningen Dissertations in Linguistics 20).
- King, J., & Kutas, M. (1995). Who did what when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7, 376–395.
- Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5, 196–214.
- 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 vice-rector for the Basque Language, available at http://www.ei.ehu.es/p289/content/eu/contenidos/informacion/grammar_euskara/en_doc/index.html
- Laka, I. 2006. “On the nature of case in Basque: structural or inherent?” en Hans Broekhuis, Norbert Corver, Jan Koster, Riny Huybregts and Ursula Kleinhenz & Jan Koster (eds.), *Organizing Grammar: Linguistic Studies in Honor of Henk van Riemsdijk*, Berlin/New York, Mouton de Gruyter, pp. 374-382.
- Levin (1983) *On the Nature of Ergativity*, MIT dissertations in Linguistics, Cambridge MA: MITWPL.
- Marácz, L.; & Muysken, P. (Eds.). (1989). *Configurationality: The typology of asymmetries*. Dordrecht: Foris.

- Marantz, A. (2005) Generative Linguistics within the cognitive neuroscience of language, *The Linguistic Review* 22, 429-445.
- Matzke, M., Mai, H., Nager, W., Rüsseler, J., & Münte, T. (2002). The cost of freedom: An ERP-study of non-canonical sentences. *Clinical Neurophysiology*, 113, 844–852.
- Mazuka, R., Itoh, K., & Kondo, T. (2002). Costs of scrambling in Japanese sentence processing. In M. Nakayama (Ed.), *Sentence processing in East Asian languages*. Stanford: CSLI.
- Miller, G.A., 1962. Some psychological studies of grammar. *American Psychologist* 17, 748–762.
- Miller, G.A. & Chomsky, N., 1963. Finitary models of language users. In: Luce, R.D., Bush, R., Galanter, E. (Eds.), *Handbook of Mathematical Psychology*, Vol. 2. Wiley, New York.
- Miller, G.A. & McKean, K., 1964. A chronometric study of some relations between sentences. *Quarterly Journal of Experimental Psychology* 16, 297–308.
- Miyamoto, E. T., & Takahashi, S. (2002). Sources of difficulty in the processing of scrambling in Japanese. In M. Nakayama (Ed.), *Sentence processing in East Asian languages*. Stanford: CSLI.
- Müller, H., King, J., & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research*, 5, 193–203.
- Münte, T. F., Matzke, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of Cognitive Neuroscience*, 9, 318–329.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97-113.
- Ortiz de Urbina, Jon (1989). *Parameters in the Grammar of Basque*. Dordrecht: Foris.
- Rebuschi, G. (1989) Is there a VP in Basque? in Marácz, L.; & Muysken, P. (Eds.) *Configurationality: The typology of asymmetries*. Dordrecht: Foris.
- Rizzi, L. *Relativized minimality*. (1990) Cambridge, MA: The MIT Press.
- Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: word-by-word variations of processing demands are revealed by even-related potentials. *Journal of Memory and Language*, 38, 150–176.
- Sekerina, I. (1997). The syntax and processing of Russian scrambled constructions. Unpublished Ph.D Dissertation. City University of New York: NY.
- Sekerina, I. (2003). Scrambling processing: Dependencies, complexity, and constraints. In S. Karimi (Ed.), *Word order and scrambling* UK: Blackwell, 301–324
- Schlesewsky, M., Bornkessel, I., & Frisch, S. (2003). The neurophysiological basis of word order variations in German. *Brain and Language*, 86, 116–128.
- Schwartz, B. D., & Vikner, S. (1996). The verb always leaves IP in V2 clauses. In A. Belletti & L. Rizzi (Eds.), *Parameters and functional heads* Oxford: OUP 11–62.
- Tamaoka, K., Sakai, H., Kawahara, J., & Miyaoka, Y. (2003). The effects of phraselength order and scrambling in the processing of visually presented Japanese sentences. *Journal of Psycholinguistic Research*, 32(4), 431–454.
- Wolff, S.; Schlesewsky, M; Hirotani, M; Bornkessel-Schlesewsky, I. 2008. The Neural Mechanisms of Word Order Processing revisited: Electrophysiological Evidence from Japanese. *Brain and Language*, 107, 133-157.
- Yamashita, H. (1997). The effects of word-order and case marking information on the processing of Japanese. *Journal of Psycholinguistic Research*, 26, 163–188.