Analyzing Interacting Phenomena: Word Order and Negation in Basque [1]

Joshua Crowgey and Emily M. Bender

1. Introduction

We present a case study of using grammar engineering to explore the analysis of interacting phenomena, as proposed in Bender (2008). In particular, we look at the case of Basque [eus] word order and negation and ask whether existing HPSG analyses of each of these can be adapted to work together. The development work was facilitated by open-source grammar engineering tools, including the Grammar Matrix customization system (Bender et al, 2002; 2010), the LKB(Copestake, 2002) and the [incr tsdb()] grammar profiling software (Oepen & Flickenger, 1998).

Although word order is a central concern for theoretical syntax[2], no HPSG analysis of negation has been presented which attempts to account for its attested ability to interact with word-order (Dryer, 1988; Miestamo, 2005).

As for negation, Kim (2000) examines sentential negation within the HPSG framework in a small selection of both European and Asian languages. Looking to Dahl (1979) for typology, Kim describes three types of negative marking strategies: morphological marking of negation, syntactic marking through a selected adverb, and negative auxiliary verbs. Word order is not impacted by negation in any of the languages Kim considers. Thus, on the basis of the existing literature, one might expect word order and negation to be independent (orthogonal) phenomena, whose analyses could perhaps be expected to be trivially interoperable crosslinguistically.

However, descriptive linguists have reported that negation interacts with word order in Basque (Saltarelli & Azkarate Villar, 1988; Manandise, 1988), constraining the range of possibilities. Thus word order and negation can not be treated entirely independently in Basque. On the other hand, we find that our independently motivated analysis of the word order facts of non-negated sentences neatly sets up the machinery needed to handle the additional constraints that arise under negation.

2. Basque

Basque is a language isolate spoken across the Western Pyrenees in Northern Spain and Southern France. It is an ergative-absolutive language with a rich system of agreement markers expressed on the finite element of verbal clauses. Most lexical verbs in Basque are incompatible with the morphological categories that indicate finiteness. For this reason, most Basque sentences contain an auxiliary verb which supports tense and mood markers, as well as agreement with the person and number of the verbal arguments. Thus a typical intransitive clause in Basque contains at least three elements: the subject, the lexical verb, and the finite auxiliary. An example is given in (1) (Manandise, 1988, 8). This example also contains what is often considered the basic order for Basque clauses (Saltarelli & Azkarate Villar, 1988).

(1) Miren ibilli da
Mary.ABS walk.PERF 3.SG.ABS.PRES
Mary has walked. [eus]

With respect to the nearly free permutations of major constituent order, Laka points out that while there is much variation, the variants are not informationally equivalent (1996). The position to the left of the lexical verb is singled out in Basque descriptions as the *galdegaia*, the object of inquiry, or the focus position. The importance of this notion is best illustrated with an example (2) (next page, Manandise, 1988, 8-9). While all of the sentences in (2) are generally grammatical, only (2b) is an acceptable answer to the

[1]We would like to thank Antske Fokkens, Esmerelda Manandise, and three anonymous reviewers for helpful discussions, scholarship, and comments. All remaining faults are our own.

[2]At least those versions of syntax which claim to be surface-oriented.

(2) a. Liburu bat nork irakurri du?
book one.ABS.SG who.ERG.SG.FOC read.PERF 3.SG.O.PRES.3.SG.A
Who has read one book? [eus]

b. Liburu bat Mirenek irakurri du.
book one.ABS.SG Mary.ERG.SG.FOC read.PERF 3.SG.O.PRES.3.SG.A
Mary has read one book. [eus]

c. Mirenek liburu bat irakurri du.
Mary.ERG.SG book one.ABS.SG.FOC read.PERF 3.SG.O.PRES.3.SG.A
Mary has read one book. [eus]

question in (2a). In the final section of this paper, we discuss the focus position's interaction with the interpretation of negation.

3. Analysis: Word order

While the ordering of major constituents in Basque is generally free, or more accurately, pragmatically determined, at least one author claims that Basque does not freely permute all combinations of the major constituents. Manandise's (1988, 15) constraint on possible orderings, is reproduced as (3).

(3) If the lexical verb is to the left of the auxiliary, then the lexical verb must be left-adjacent to the auxiliary.

(4) *Liburu irakurri Mirenek du.
book.ABS.SG READ.PERF Mary.ERG.SG AUX
Mary has read a book. [eus]

Manandise further claims that this constraint holds for Basque main clauses with up to three NPs and that beyond this constraint, no further checks on major constituent order apply. The sentence in (4), for example, is ruled out by (3).

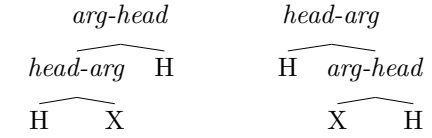
Manandise's constraint suggests a bifurcation of the data into those sentences in which the auxiliary precedes the lexical verb and those in which it follows. The regular expressions in (5) schematize these two (complementary) patterns. In aux-first strings, the NPs can occur freely around and between the auxiliary and the verb, as summarized in (5a). When the verb precedes the auxiliary, however, NPs may not intervene between them, as shown in (5b). First we turn our attention to achieving free word ordering amongst the first group.

(5) a. /NP* Aux NP* V NP*/
b. /NP* V Aux NP*/

For the aux-first sentences (5a), we wish to allow free word order. We begin with the default analysis for free word order from the Grammar Matrix customization system (Fokkens, 2010). As Fokkens notes, handling free word order entails much more than allowing unconstrained syntax. In addition to licensing all of the orders, the syntactic arguments need to be linked to the correct semantic positions. Fokkens handles this with a series of binary-branching rules of the familiar head-nexus types.[3] However, simply providing both head-final and head-initial rule types for each phrasal rule leads to spurious ambiguity. Consider the example in (6), if a *head-arg* rule has both head-final and head-initial forms, then both of these trees will be valid parses for the string *H X H* with no semantic difference between them.

[3]This analysis, like all analyses provided by the Grammar Matrix customization system, uses binary-branching rules. This is due in part to technical considerations of the grammar implementation language: systems that interpret tdl require phrase structure rules to have fixed arity and fixed order of daughters. Since we need to know which daughter is which in order to correctly constrain e.g., agreement (not to mention semantic composition), a grammar with binary branching rules needs far fewer rules than one that strives for flatter structures.

(6)



Fokkens' analysis constrains the space of possible analyses by requiring the grammar to apply any head-initial rules before any head-final rules.[4] In this way, left and right branching rules cannot factor across each other in the parse forest. Instead, given a [Aux, NP, Verb] sequence, only the bracketing [[Aux NP] Verb] is licensed.

The grammar must also rule out spurious ambiguity for sequences of the type [Aux, Verb, NP]. There is potential here for two parses using only head-initial rule types. The grammar we have designed enforces a single bracketing of these sequences automatically by taking advantage of the need for argument agreement on the auxiliary.

Auxiliaries in Basque agree with up to two arguments of the clause. We model this in the grammar with argument composition (Hinrichs & Nakazawa, 1990), and then simply having the inflected auxiliaries constrain the agreement features of all NP arguments on their valence lists. The feature structure in (7) shows some of the constraints stipulated on an auxiliary lexical type. This type inherits from core grammar type *arg-comp-aux-no-pred* (Bender et al., 2002). Note the nonempty specification for the auxiliary's first complement's first complement.

(7)
$$\begin{bmatrix} \textit{transitive-abssg-aux-lex} \\ \text{SUBJ} \left\langle \left[\text{CASE } \textit{erg} \right] \right\rangle \\ \text{COMPS} \left\langle \begin{bmatrix} \text{FORM } \textit{nonfinite} \\ \text{COMPS} \left\langle \left[\text{CASE } \textit{abs} \right] \right\rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

We leverage this nonempty specification, along with the fact that in typical in HPSG grammars head-argument rules cancel elements off the valence list as the head path is projected, to constrain the analysis of sequences of the form [Aux, Verb, NP]. If the verb first combines with its complement a VP (COMPS satisfied) structure is the result. This VP is incompatible with the specification on the auxiliary's complement (as in (7)). The only licensed bracketing then, is [[Aux Verb] NP].

This analysis of the first set of data allows us to capture the free word order properties of Basque while avoiding spurious ambiguity. Let us now turn to the set of examples in which the lexical verb precedes the auxiliary.

Our analyses of the orders schematized in (5b) can't simply be the mirror image of those in (5a), because we need to rule out any strings in which an NP intervenes between the verb and the auxiliary. To accomplish this, the grammar is augmented with a verbal complex analysis. This option is also a part of the word-order library (Fokkens, 2010) that the Grammar Matrix customization system makes available. Rather than making the verbal complex available for all sentences, we use it only for the class of sentences schematized in (5b).

The grammar's verbal complex rule is presented in (8). This rule-type inherits from both *basic-head-1st-comp-phrase* and *head-final* types (Bender et al., 2002), which implement the Valence Principle and head-finality, respectively.

(8)
$$\begin{bmatrix} \textit{comp-aux-phrase} \\ \text{SYNSEM|LOCAL|CAT|HEAD} \begin{bmatrix} \textit{verb} \\ \text{AUX} \quad + \end{bmatrix} \\ \text{NON-HEAD-DTR|SYNSEM|LOCAL|CAT|HEAD} \left[\textit{verb} \right] \\ \text{HEAD-DTR|SYNSEM|LIGHT} + \end{bmatrix}$$

The feature, [VC *luk*] (mnemonic for verbal cluster), is defined in the grammar on phrasal and lexical *synsems*.[5,6] Lexical verb types are stipulated as [VC +], while auxiliaries are set to [VC −]. Head-complement rule types are then defined to inherit their VC value from their non-head daughter. In this way, an auxiliary which has picked up its lexical verb complement will form a phrase which is [VC +]. The value of VC on a phrase thus indicates whether or not the lexical verb is present in that phrase.

[4] A feature ATTACH and a small value hierarchy are employed to effect this. See Fokkens (2010) for details.

[5] *lex-rule* types are also annotated such that they pass up the value of VC through the inflectional pipeline.

[6] Named after Polish logician Jan Lukasiewicz, *luk* is a generalization of the type *bool* that is consistent with three values: $\{+, -, na\}$.

To see how these types rule out phrases which contain one or more NPs intervening between the lexical and auxiliary verbs, first consider given the sequence [Verb, NP, Aux]. If the lexical verb first picks up the argument, the resulting valence list is shortened and the auxiliary will not be able to access (or constrain) case information on the NP (as described above). In this way, the bracketing [[Verb NP] Aux] is ruled out. Secondly, we specify that in comp-head and subj-head rules, the head daughter must be [VC +]. In this way we avoid the bracketing [Verb [NP Aux]]. These two aspects of the grammar thus rule out the sequence under consideration, and it should be clear that the same facts generalize to cases with more than a single intervening NP; sequences that match this regular expression: /Verb NP^+ Aux/, are equally unparseable.

Again, we confront the potential for spurious ambiguity on sequences of the form [Verb, Aux, NP]: we do not wish to allow both bracketings [[Verb Aux] NP] and [Verb [Aux NP]]. The verbal complex rule does not inherit from the *head-final-head-nexus* type which enforces that head-initial rules apply before head-final ones. This is because we use the verbal complex rule to ensure that the Verb and Aux elements appear adjacent to each other and despite the fact that the Aux element heads the phrase, we want the verbal complex rule to apply before any argument attachment in any licensed parse of the verb-first data. This is the motivation for the stipulation [LIGHT +] in the *comp-aux-phrase* presented in (8). The feature LIGHT is defined on *synsem*s with a value *luk*. Lexical items are [LIGHT +], while phrases are [LIGHT −]. This stipulation ensures that the verbal complex rule applies before the auxiliary picks up any arguments in any successful parse.

The grammar as we have defined it thus provides an implementation of Manandise's constraint on word order—modeling the partially free word order observed in Basque in an explicit, testable form. The next section discusses the overlay of the negation analysis onto the grammar presented.

4. Negation

Sentential negation in Basque is accomplished by the prefixation of a negative morpheme, *ez*, to the finite element (Manandise, 1988, 12; Saltarelli & Azkarate Villar, 1988, 92). Manandise does not discuss the bound or free status of this morpheme, but she does present examples without whitespace between *ez* and the auxiliary—flouting typical orthography conventions—in her introductory exposition. Saltarelli, on the other hand, explicitly calls this morpheme a particle, entailing an analysis as a free morpheme, but does not offer any argument. We follow Manandise here in treating negation as bound for reasons analogous to those given in Kim (2000, 34) for the Korean morpheme *an*. Both Basque and Korean allow free permutation of syntactic elements (almost free in Basque), but the position of *ez* is fixed to the auxiliary verb. There is no possible intervention of adverbials. These facts would have to be dealt with in the syntax if we treat *ez* as free, by treating it as bound, the Grammar Matrix's implementation of the Lexical Integrity Principle (Bresnan & Mchombo, 1995; Kim, 2000) ensures that bound morphemes cannot stray from their hosts. In our analysis, *ez* is added to Aux types by a lexical rule.

As mentioned in the introduction, negation interacts with word order in Basque. The interaction is such that although Basque allows main clauses in which the lexical verb appears to the right or to the left of the auxiliary verb, under negation, only those constructions in which the main verb follows the auxiliary verb are licit.[7] The overlaying of this constraint upon Manandise's filter (3) means that under negation we find a narrowing of possible word orders; only those sentences described by the expression in (5a) are compatible with negation, as shown in

[7]This is only true of main clauses. In subordinate clauses, the lexical verb precedes the finite element because of an independent constraint on subordinate clauses which requires that the finite element appear finally. While the solution may rely on additional specialized rules, we believe that the approach presented here will scale as we extend our fragment to handle subordinate clauses as well.

(9):

(9) /NP* ez-Aux NP* V NP*/

If we were to assume that negation and word order are independent—and just add the lexical rule to attach the negative morpheme to auxiliary verbs—the grammar will overgenerate, licensing strings that match the pattern in (10), even though these are uniformly ungrammatical:

(10) */NP* V ez-Aux NP*/

The analysis of word-order given above required the introduction of a construction-specific rule—a verbal complex rule which combined a left-adjacent lexical verb with a selecting auxiliary. We engineered this rule in such a way that it bisects *a priori* possible sentences into two groups: aux leading (5a) vs verb leading (5b). The verbal complex rule only and always appears in successful parses of the verb-leading examples. Thus, it provides a natural target for constraints that should apply to only one group or the other. We implement the constraint via a flag feature whose value is set by the negation rule and we stipulate an incompatible value for the instances of the verbal complex rule.

The grammar presented here thus defines [NEGATED *luk*] on *synesm*s. We modify the lexical rule that carries out negation such that it is [NEGATED +]. In this way, the feature NEGATED encodes whether or not a verb has been negated. Finally, we add to definition of the *comp-aux-phrase* (verbal complex rule) the stipulation [NEGATED −]. The interaction of these components conspires to rule out any examples in which the lexical verb appears to the left of a negated auxiliary.

5. Conclusion and Outlook

We have seen that the existing analyses of (mostly) free word order and negation can in fact be adapted to work together to capture the facts of Basque. A key property of this success was the constructional approach taken by the word-order analysis, which led to the availability of a specific rule on which to hang the constraints about negation.

The next step in this work is to consider the interaction of both word order and negation with focus. Focus is encoded in Basque word order, but negation also interacts with the focus position in Basque. In Basque, the element which appears just to the left of the lexical verb is focused. When this element is the negating auxiliary, Manandise (1988) treats the negation as having sentential scope. When the focused element is a NP, Manandise treats this construction as constituent negation.

These issues concern the interfaces between information structure, syntax and semantics, which we contend can only be fully understood via modeling with a precise, machine-readable grammar. We believe that the analyses presented here will form the basis of a grammar that can be extended to cover interactions with additional phenomena, including focus.

References

Bender, E. M. (2008). Grammar engineering for linguistic hypothesis testing. In N. Galord, A. Palmer, & E. Ponvert (Eds.), *Proc. TLSX* (pp. 16–36). CSLI.

Bender, E. M., Drellishak, S., Fokkens, A., Goodman, M. W., Mills, D. P., Poulson, L., et al. (2010). Grammar prototyping and testing with the LinGO grammar matrix customization system. In *Proc. ACL 2010 system demonstrations* (pp. 1–6).

Bender, E. M., Flickinger, D., & Oepen, S. (2002). The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In J. Carroll, N. Oostdijk, & R. Sutcliffe (Eds.), *Proc. workshop on grammar engineering at COLING 2002* (pp. 8–14).

Bresnan, J., & Mchombo, S. (1995). The lexical integrity principle: Evidence from Bantu. *Natural Language & Linguistic Theory*, *13*(2), 181–254.

Copestake, A. (2002). *Implementing Typed Feature Structure Grammars. 2002.* CSLI.

Dahl, Ö. (1979). Typology of sentence negation. *Linguistics*, *17*(1-2), 79–106.

Dryer, M. (1988). Universals of negative position. *Studies in syntactic typology*, 93–124.

Fokkens, A. (2010). *Documentation for the Grammar Matrix word order library* (Tech. Rep.). Saarland University.

Hinrichs, E., & Nakazawa, T. (1990). Subcategorization and VP structure in German. In S. Hughes & J. Salmons (Eds.), *Proc. 3rd symposium on Germanic linguistics.* Benjamins.

Kim, J. (2000). *The grammar of negation: A constraint-based approach.* CSLI Publications Stanford.

Laka, I. (1996). *A brief grammar of euskara, the basque language.* http://bit.ly/iimVwh.

Manandise, E. (1988). *Evidence from Basque for a new theory of grammar.* Garland.

Miestamo, M. (2005). *Standard negation: The negation of declarative verbal main clauses in a typological perspective.* Walter de Gruyter.

Oepen, S., & Flickenger, D. P. (1998). Towards systematic grammar profiling: test suite technology ten years after. *Journal of Computer Speech and Language*, *12 (4)*, 411 – 436.

Saltarelli, M., & Azkarate Villar, M. (1988). Basque. *Croom Helm*.