An outline of a Basque computerized grammar

Holmer, Arthur; Sigurd, Bengt

2001

Document Version:
Other version

Lund University, Dept. of Linguistics
Working Papers 48 (2001), 115–133

An outline of a Basque computerized grammar

Arthur Holmer and Bengt Sigurd

Introduction

Basque is spoken by approximately 660,000 people (Trask 1997) in the Basque Country, straddling the border between northern Spain and southern France, particularly in rural areas of the Bizkaia and Gipuzkoa coastline, as well as inland Gipuzkoa and the foothills of the Pyrenees in Navarre. Virtually all Basque speakers are bilingual with either Spanish or French. Nevertheless, Basque is, due to its official status in the Basque autonomous region in northern Spain, no longer as seriously endangered as previously.

Basque has several dialects which differ greatly from one another. The present paper deals with the standard version, which is, together with Spanish, the official language of the Basque autonomous region. This variety is commonly termed *euskara batua* (lit. 'Unified Basque').

The genetic origin of Basque is unknown. Its only known relative is Aquitanian, which was spoken in a large area of south-western Gaul during the time of the Roman occupation (Trask 1997), and of which fragmentary inscriptions have survived. Other than that, no genetic relationship between Basque and any other language has been proven. Typologically, Basque shares certain features with Caucasian languages in particular, but these common features are not necessarily evidence of genetic relationship.

Basque is a morphologically ergative language with accusative syntax. Morphological ergativity implies that the subject of a transitive verb is realized in ergative case, as opposed to the object of a transitive verb and the subject of an intransitive verb, which are realized in absolutive. Syntactic accusativity implies that the ergative noun phrase behaves like a clause subject according to standard diagnostic tests such as control, clausal coreference and reflexive binding (cf. Ortiz de Urbina 1989, Holmer 1999). Neutral word order is SOV, but other orders are also common. It has postpositions (both suffixal and free), and G-N word order within the noun phrase. A relative clause precedes its

head. However, unusually for an SOV language, it has N-Adj word order within the noun phrase. Basque has triple verb agreement (subject, object and indirect object; indirect object agreement is optional in northern dialects of Basque).

The (relatively) free word order and the complex inflectional system offer interesting challenges to grammatical models. This paper presents a computerized formal grammar of Basque and shows how the free word order and the complicated agreement can be handled by a computerized formal grammar called Permutational Grammar (PG; cf Eeg-Olofsson & Sigurd 2001). This grammar is a variant of SWETRA (Referent) Grammar which has been presented in various publications, notably Sigurd (ed.), *Computational grammars for analysis and machine translation* (1994).

We will present the characteristics of Permutational Grammar at the same time as the different Basque grammatical patterns are outlined. Some basic morphological rules are also presented, but only a few lexical items.

The Permutational Grammar of Basque can be used in automatic translation given equivalent grammars. A Basque-Chinese translation example will also be demonstrated.

Some preliminary notes on Basque syntax patterns

The following are some examples of possible word orders in a simple Basque sentence. Possible positions where an adverbial may be inserted are illustrated by the position of the adverb *gaur* 'today'.

1. a. Peruk (gaur) Amaia (gaur) ikusi du (gaur).
   Peru-ERG (today) Amaia-ABS (today) seen has (today)
   'Peru has seen Amaia today.'

   b Peruk (gaur) ikusi du (gaur) Amaia (gaur).

   c. Amaia (gaur) Peruk (gaur) ikusi du (gaur).

   d. Amaia (gaur) ikusi du (gaur) Peruk (gaur).

   e. (gaur) ikusi du (gaur) Peruk Amaia (gaur).

   f. (gaur) ikusi du (gaur) Amaia Peruk (gaur).

In all the above examples *Peruk*, the ergative (ERG) form of the proper name *Peru* is the subject. *Amaia* which is in absolutive Case (henceforth ABS) is the object. The word order changes thus have no effect on the functional interpretation of the noun phrases.

While the first pattern (with SOV word order) is the most common, all the others are acceptable, given a suitable context. The most important factor

governing Basque word order is information structure – thus the focussed element (such as a *wh*-word in a *wh*-question) is obligatorily placed immediately before the main verb.

Another constraint concerns the relative ordering of main verb and auxiliary. In affirmative clauses, the main verb precedes the auxiliary as in *ikusi du* [seen has3sA-3sE] 'has seen'. In negated clauses, on the other hand, the auxiliary obligatorily occurs immediately after the negator *ez*, whereas the main verb is positioned to the right of this cluster (not necessarily immediately).

The subject of an intransitive clause is realized in ABS case (the same case as that of an object of a transitive construction). If there is an auxiliary in the sentence it is inflected for person and number agreement with the subject, direct object and indirect object (if these arguments are present in the argument structure of the verb).

The distinction between transitive and intransitive verbs is partly semantic (depending on the argument structure of the verb) and partly grammatical (depending on the choice of auxiliary and the case-marking of the subject). There is a mismatch between these two criteria, however, and there is a set of verbs which are semantically intransitive but behave grammatically as transitives (with ERG case-marking of the subject and transitive auxiliary selection). Example (2) illustrates such a verb.

2. Neska-k dantzatu du.
girl-ERG dance has-3sA-3sE
'The girl has danced.'

In generative models, such verbs are referred to as unergative verbs. For the purpose of this paper, the term *semi-transitive* will be used. The analysis of this set of verbs is not crucial for the present analysis, for a detailed discussion the reader is referred to Holmer 1999.

While most verbs in Basque are inflected using an auxiliary verb (analytic inflection), some verbs are capable in their own right of bearing tense-marking and person agreement. This type of inflection is referred to as synthetic inflection. Synthetic inflection is only open to a handful of verbs, and even for these analytic inflection is usually an option. Furthermore, the entire range of moods and tenses available for analytic inflection is greatly restricted with synthetic inflection, so all synthetically inflected verbs must, in certain moods, be inflected analytically instead. However, the present paper is solely concerned with indicative mood, so this problem will not be addressed here. Synthetic inflection is exemplified by (3).

3. Gizon-ek eskutitza-k Amaia-ri darama-zki-o-te.
man-ERG.PL letter-ABS.PL Amaia-DAT bring-3pA-3sD-3pE
'The men bring the letters to Amaia.'

ABS is realized by a Ø-suffix, ERG by *-k* and DAT by *-ri*. The plural marker is *-k* in ABS case.

Since Basque displays multiple verb agreement, a typical verb paradigm involves a two-dimensional table rather than a simple list. This is illustrated below for the present tense of the transitive auxiliary 'to have' (4). Further verb tables are required for various modes and tenses, as well as for ditransitive verbs, creating verbal systems of enormous complexity. In the present version, we are primarily concerned with sentence syntax, so we shall restrict ourselves to a subset of these forms, listing them lexically rather than deriving them from a separate morphological component. A more detailed morphological treatment is given in Holmer & Sigurd (fc).

4. Inflection of 'to have' (PRES)

ERG		SG				PL		
ABS		1	2	2pol	3	1	2	3
SG	1	–	nauk/-n	nauzu	nau	–	nauzue	naute
	2	haut	–	–	hau	haugu	–	haute
	2pol	zaitut	–	–	zaitu	zaitugu	–	zaituzte
	3	dut	duk/-n	duzu	du	dugu	duzue	dute
PL	1	–	gaituk/-n	gaituzu	gaitu	–	gaituzue	gaituzte
	2	zaizuztet	–	–	zaituzte	zaituztegu	–	zaituztete
	3	ditut	dituk/-n	dituzu	ditu	ditugu	dituzue	dituzte

Another interesting feature of Basque, in keeping with the head-final nature of the language, is its placement of the relative clauses in front of its head (the correlate), cf. examples (5a-c). The relative marker is formed by suffixing the relativizer *-en* to the auxiliary verb.

5. a. Peru ikusi du-en neska joan da.
Peru see have-3sA-3sE.PRES-REL girl go is-3sA
'The girl who has seen Peru has gone.'

b. Peru ikusi zuen neska joan da.
Peru see have-3sA-3sE.PRET-(REL) girl go is-3sA
'The girl who saw Peru has gone.'

c. Peru-k ikusi du-en neska joan da.
Peru-ERG see have-3sA-3sE.PRES-REL girl go is-3sA
'The girl whom Peru has seen has gone.'

We will present the formal rules for these sentences and a number of others below. However, we will not discuss all the complexities and subtleties of Basque syntax in detail. For a closer analysis, the reader is referred to the references given in the bibliography.

Permutation Grammar rule for transitive sentences

Basque word order patterns have to be distinguished on the basis of the argument structure of the verb and the mode of the sentence. Negated sentences also have to be treated separately. Furthermore, patterns with synthetic verb inflection have to be distinguished from cases with analytic inflection. Let us start with a transitive sentence with the object in focus (6).

6.
Peru-k	gaur	Amaia	ikusi	du.
Peru-ERG	today	Amaia-ABS	seen	have-3sA-3sE

'Peru has seen Amaia today.'

The Permutational Grammar rules are generative rewriting (arrow) rules which can be run directly in the programming language Prolog (using the formalism Definite Clause Grammar, DCG). The rule below states that there is a Basque sentence type (`bs`), which covers declarative sentences (Mode:`d`(eclarative)), and in which the constituent immediately before the verb is in focus (`Fo`).

7.
```
bs(d,Fo,[subj(N1),pred([V,As,T]),obj(N2),advl(A1),
advl(A2),advl(A3),co(N1,Mo,C,F2)]) -->
{permute([bnp(Agr1,N1),badvp(A1),bnp(Agr2,N2),
badvp(A2),bvt([V,As]),bauxt(Agr1,Agr2,T),badvp(A3),
btail(C,F2)],M)},
  {imbefore(bvt([V,As]),bauxt(Agr1,Agr2,T),M)},%v before aux
{imbefore(Fo,bvt([V,As]),M)}, % focus before main verb
{Fo≠badvp([])},{last(btail(C,F2),M)}, % tail must be last
surf(M).
```

The functional representation of this sentence type is what is included in the bracket immediately to the left of the arrow, i.e. (8).

8.
```
[subj(N1),pred([V,As,T]),obj(N2),advl(A1),advl(A2),
advl(A3),co(N1,Mo,C,F2)].
```

This functional representation shows the meaning (value) of the functional roles when all rules including phrase rules, morphological rules and lexical rules are applied and the variables are instantiated. Capital letters denote variables in Prolog. The meaning (value) of the subject is denoted by the variable `N1` and when instatitiated it may be `m(peru,prop)`, which is the semantic representation of the given name *Peru*, regardless of whether it be

realized in ergative, absolutive or dative case. The value of the predicate is shown by `[V,As,T]` – the main verb value (`V`), the Aspect value (`As`) and the tense value (`T`), the latter two of which are given by the auxiliary. When instantitiated by a main verb form and an auxiliary it may be realized as the following (9).

9. `[see,perf,pres].`

The bracket including the functional roles (8) includes three adverbial slots, which is generally enough for ordinary sentences. There is also a last constituent named `co`. This constituent (which we will not treat in detail here) has to do with possible coordination and the use of full stop or question mark in other cases. Its simplified surface equivalent is `btail`. The category labels should otherwise be self-explanatory.

The realization of the functional representation is shown to the right of the generative arrow in another bracket parenthesis where we can see e.g. that the `bnp(Agr,N1)` corresponds to the subject (`N1`) in the functional representation. The Basque categories have been given the prefix `b` to distinguish them from English categories which have the prefix `e` when several grammars are run simultaneously as in machine translation. A possible adverbial `badvp(A1)` is placed next in the surface list of phrases and then another noun phrase `bnp(Agr2,N2)`, corresponding to the object (`N2`) in the functional representation. The variable `N1` is the value of the subject as given by the lexical words and the agreement variables `Agr1` and `Agr2` cover complex values of number, person and case. If a noun phrase has the variable `Agr1` and the verb also carries the variable `Agr1` this means that the phrases have the same agreement features, or technically more correctly: features which can be unified. The checking of agreement is handled by the unification mechanism in Prolog. By this different markers (e.g. `erg` and `abs`) cannot unify, but identical markers can. If one (or both) markers are uninstantiated unification is also possible. Uninstantiated values are written as underscore (_). Thus `agr(sg,_,erg)` and `agr(_,_,erg)` can unify, but `agr(sg,_,abs)` and `agr(pl,_,abs)` cannot. Uninstantiated values get numbers when the computer program is run. The middle position in the complex agreement variable is reserved for the category of person, which we do not specify in this version.

The square bracket parenthesis after the arrow shows one basic order, but the preceding Prolog predicate `permute` allows this order to be changed in all possible ways (permutations) in order to generate other word orders. It is

the use of the permutation predicate which gives this variant of Swetra grammar its great power. We will not define the Prolog predicate `permute(L,L1)` here.

The following are some lexical items including Basque transitive (`bvt`) verbs.

10.
```
bvt([see,perf]) --> [ikusi].
bvt([see,imp]) --> [ikusten].
bvt([see,fut]) --> [ikusiko].
bvt([learn,perf]) --> [ikasi].
bvt([learn,imp]) --> [ikasten].
bvt([learn,fut]) --> [ikasiko].
bvt([write,perf]) --> [idatzi].
bvt([write,imp]) --> [idatzen].
bvt([write,fut]) --> [idatziko].
```

The following are conjugated auxiliaries are listed as lexical entries.

11.
```
bauxt(agr(sg,_,erg),agr(sg,_,abs),pres) --> [du].
bauxt(agr(pl,_,erg),agr(sg,_,abs),pres) --> [dute].
bauxt(agr(sg,_,erg),agr(pl,_,abs),pres) --> [ditu].
bauxt(agr(pl,_,erg),agr(pl,_,abs),pres) --> [dituzte].
bauxt(agr(sg,_,erg),agr(sg,_,abs),past) --> [zuen].
bauxt(agr(pl,_,erg),agr(sg,_,abs),past) --> [zuten].
bauxt(agr(sg,_,erg),agr(pl,_,abs),past) --> [zituen].
bauxt(agr(pl,_,erg),agr(pl,_,abs),past) --> [zituzten].
```

Word order constraints

If no constraint was introduced on the sequence of phrases in the list M, all logically possible permutations would be permitted. However, as mentioned, it is a characteristic of Basque that the main verb must occur immediately before the auxiliary (except in negated sentences) and this condition is implemented by the code within the curly condition brackets used in Prolog.

12. `{imbefore(bvt([V,As]),bauxt(Agr1,Agr2,T),M)}`

This code states that the constituent `bvt([V,As])` occurring in any of the permuted lists (`M`) must occur immediately before the auxiliary constituent `bauxt(Agr1,Agr2,T)`. Note that it is the auxiliary that carries the agreement features `Agr1` (also to be found in the first NP representing the subject) and `Agr2` also to be found in the NP representing the object. The main verb carries the main meaning (`V`) and the aspectual meaning denoted by `As`. The auxiliary contributes the tense meaning denoted by the variable `T`.

A futher condition is implemented by the next curly brackets in the rule. It states that the focus constituent represented by the variable (`Fo`) must occur before the verb in all permutations.

13. `{imbefore(Fo,bvt([V,As]),M)}`

One other linearity condition is needed, namely the requirement that the full stop or a following coordinated sentence be placed last. This condition is achieved by the Prolog predicate `last` .

The variables in `btail` have to do with the situation when there is a coordinated sentence and not just a full stop. `C` covers the coordinator, whereas `F2` is the functional representation of a coordinated clause. We will not give more details here.

14. `{last(btail(C,F2),M)}`

The final predicate `surf(M)` instructs the generator to realize the permuted list (`M`) as a linear string of words at the surface.

The syntactic Prolog rule above can generate examples such as the following which are all acceptable in Basque. The translations given are only some of the possible equivalents (15).

15. a. Peruk gaur Amaia ikusi du. (SAOV)
‘Today Peru has seen Amaia.’

b. Peruk hemen Amaia ikusi du gaur. (SA_1OVA_2)
‘Here Peru has seen Amaia today.’

c. Gaur Amaia ikusi du Peruk hemen. (A_1OVSA_2)
‘Today Peru has seen Amaia here.’

d. Gaur Amaia Peruk ikusi du. (AOSV)
‘Today Peru has seen Amaia.’

The use of the predicate `imbefore` excludes ungrammatical examples such as (16).

16. a. *Amaia Peruk ikusi gaur du.
Amaia Peru-ERG see today have-3sA-3sE
(*V-AUX not adjacent)

b. *Peruk ikusi Amaia du.
Peru-ERG see Amaia have-3sA-3sE
(*V-AUX not adjacent)

c. *Peruk du ikusi Amaia
Peru-ERG have-3sA-3sE see Amaia
(*AUX before V)

One may test the grammar by parsing a Basque sentence. This is shown by the following call for the analysis of `[gaur, 'Peruk', 'Amaia',`

`ikusi, du, hemen]`, where we get the solution that the mode is `d` (declarative), the object is in focus, and the functional representation is as shown in F (17).

17.
```
bs(M, Fo, F, [gaur, 'Peruk', 'Amaia', ikusi, du, hemen],[])
No.1 : M = d, Fo = bnp(agr(sg, _92105, abs),
m(amaia, prop)), F = [subj(m(peru, prop)),
pred([see, perf, pres]), obj(m(amaia, prop)),
advl(m(today,    _94197)), advl([]), advl(m(here, _93507)),
co(m(peru, prop), _10191, [.], [])]
```

Alternatively one may ask the grammar to produce a sentence given certain facts: the mode required, what is to be the focus, what is to be the subject, the predicate, etc. The following call shows how we ask for a sentence with *Peru* as subject, *girl* as object and the adverb meaning *here* in sentence focus. The result of the call is given in (18). The morphological analysis follows.

18.
```
bs(d, badvp(m(here,_)),[subj(m(peru, prop)),
pred([see, perf, pres]), obj(m(girl, sg)), advl([]),
advl(m(here,_)), advl([]), co(m(peru, prop),
_46128, [.], [])], X, [])
X = [ 'Peruk',neska, hemen,ikusi, du]
```

Peru-k	neska	hemen	ikusi	du.
Peru-ERG	girl-(DET-ABS)	here	seen	3sA-3sE-AUX

‘Peru has seen the girl here.’

The grammar also requires lexical entries, which for an adverbial phrase may take the form as in (19).

19.
```
badvp([]) --> []. % empty, no adverb
badvp(m(today,_)) --> [gaur].
badvp(m(here,_)) --> [hemen].
```

The first item in the list gives an empty list indicating an optional adverbial. The syntactic rule given above (7) allows three adverbial positions which may then be empty.

Programming noun phrases

The syntactic rules use the category noun phrase `bnp`, representing the various functional roles subject, object and dative object. The noun phrase carries different agreement values which fit the requirements of the verb.

Noun phrases may include relative clauses, determiners, adjectives and numbers. In order to handle these cases, successively smaller noun phrases of different hierarchical order (bar values) are used (`bnp3, bnp2, bnp1`).

The rule in (20) is rather trivial in that it indicates that a bnp may consist simply of a phrase of level 3: bnp3. (A further alternative rule defines another possible bnp, namely a relative clause followed by a bnp3, and the purpose of this is to allow for noun phrases with relative clauses). Thus, this rule simply allows for a bnp which does not contain a relative clause.

20.
```
bnp(Agr,N) --> bnp3(Agr,N). % only np3 (no rel clause)
```

The rule in (21) allows the formation of a bnp3 which consists simply of a lexical head bnlex, provided that this lexical head is a proper noun (prop). The consequence of this is that a proper noun such as a name may occur as a bnp in its own right, or may occur together with a relative clause (by virtue of forming a bnp3 which in turn may be combined with a relative clause to form a bnp.

21.
```
bnp3(Agr,N) --> bnlex(Agr,N), {N=m(_,prop)}.
```

Examples of lexical entries of proper nouns are given in (22).

22.
```
bnlex(agr(sg,_,abs),m(amaia,prop)) --> ['Amaia'].
bnlex(agr(sg,_,abs),m(peru,prop)) --> ['Peru'].
```

Articles

In Basque the definite marker *(a)* is suffixed to the noun phrase (if an NP contains several words, the articles is suffixed to the last word). The following rule displays the affixation of the article to an NP containing only one word. The rule (23) is the first step in our derivation of nominal morphology. It affixes the definite determiner *-a* to a bnp2 to create a bnp3, e.g. *gizona* 'the man' from *gizon* 'man', *ardoa* 'the wine', from *ardo* 'wine'.

23.
```
bnp3(Agr,[the,B],[F1]) :-
bnp2(Agr,B,[F],[]),B=m(B1,P),P≠prop,
(ccat(R,a,F),F1=F ; % don't add if ending in a
ccat(F,a,F1)). % add def art a
```

The operator :− denotes implication in Prolog in such a way that what comes before the :− operator is true if what comes after the :− is true. This rule can be verbalized such that if there is a bnp2 with the agreement features Agr, the meaning B and the form F and it is not a proper name (P≠prop), a corresponding bnp3 may be formed by concatenating (ccat – a variant of the built-in predicate concat) an *a* to the form F, giving a new form F1.

The first section of the condition (ccat(R,a,F), F1=F) states that if the form F already ends in *-a*, no extra suffix is added (F1 = F). The marker

the, meaning definite, is added to the meaning B as seen in the square brackets [the,B]. If there are several words in F, e.g. if there is a following adjective, the *-a* has to be concatenated to the last item. We will not present this variant of the rule here.

A further alternative to forming a bnp3 is the addition of the indefinite article *bat* 'one'. This is illustrated in (24). The ensuing meaning representation includes the indefinite article *a* (from English). Note that this is only possible if bnp2 is not a proper name (P≠prop). We have not listed the definite nor the indefinite article in Basque as separate categories.

24.
```
bnp3(Agr,[a,B]) -->
bnp2(Agr,B), {B=m(B1,P),P≠prop},[bat].
```

Demonstratives and adjectives

As an alternative to affixing of the definite article *-a*, a demonstrative may be added to a bnp2 to yield a bnp3. Such elements include the demonstratives *hau* 'this' and *hori* 'that'. Example (25) illustrates a rule which derives bnp3 from bnp2 followed by a demonstrative (bdem). The lexical items for the two demonstratives are also given. The meaning of the demonstrative item (D) is placed before the meaning of np2 in brackets [D,B].

25.
```
bnp3(Agr,[D,B]) --> bnp2(Agr,B),bdem(D).
bdem(this) --> [hau].
bdem(that) --> [hori].
```

A grammatical bnp2 may consist of a bnp1 followed by an adjective. This case is covered by rule (26). One possible adjective, *handi* 'big', is listed in (26).

26.
```
bnp2(Agr,B) --> bnp1(Agr,B).
bnp2(Agr,[A,B]) --> bnp1(Agr,B), ba(A). % with adj
ba(big) --> [handi].
```

A bnp2 may also consist of a simple lexical item and this case is covered by the following rule, which uses lexical items as those listed above.

27.
```
bnp2(Agr,B) --> bnlex(Agr,B).
```

The following (28) is a list of lexical items with agreement features and meaning representations. Word meanings are written in the format m(L,G), where L is the lexical meaning and G the grammatical meaning (e.g. sg, pl, prop, pres, past). We will show below how the absolutive form can be used as a basis for generating the other cases.

28.
```
bnlex(agr(sg,_,abs),m(book,sg)) --> [liburu].
bnlex(agr(sg,_,abs),m(girl,sg)) --> [neska].
bnlex(agr(sg,_,abs),m(basquelanguage,_)) --> [euskara].
bnlex(agr(sg,_,abs),m(wine,_)) --> [ardo].
bnlex(agr(sg,_,abs),m(boy,sg)) --> [mutil].
bnlex(agr(sg,_,abs),m(person,sg)) --> [gizon].
bnlex(agr(sg,_,abs),m(letter,sg)) --> [eskutitza].
bnlex(agr(sg,_,abs),m(money,_)) --> [diru].
bnlex(agr(sg,_,abs),m(book,sg)) --> [liburu].
bnlex(agr(sg,_,abs),m(table,sg)) --> [mahai].
bnlex(agr(sg,_,abs),m(house,sg)) --> [etxe].
```

The above rules are sufficient to generate noun phrases up to the level of `bnp3` (i.e. noun phrases without relative clauses).

Number attributes

Alternatively, a numeral may be added to `bnp2` to create a `bnp3` . This case is given by (29). Numerals do not occur with proper names, as is stated by `P≠prop`. Some numerals are also listed. We will not treat the case when a numeral (before) cooccurs with a definite article (after) here.

29.
```
bnp3(Agr,[N,B]) -->
bnum(N),bnp2(Agr,B), {B=m(B1,P),P≠prop}. % with num
bnum(2) --> [bi].
bnum(3) --> [hiru].
bnum(4) --> [lau].
bnum(5) --> [bost].
```

Noun phrases with relative clauses

The category `bnp` may include a relative clause as is shown in the following example (30).

30. Peru ikusi duen Amaia
Peru-(ABS) seen 3sA-3sE-AUX-REL Amaia
‘Amaia who has seen Peru’.

The relative clause in Basque precedes the noun (in our system `bnp3`). The agreement features of the head should be carried over to the whole noun phrase. The meaning of the relative clause is represented by `R` in the square brackets `[N,R]`, where `N` is the meaning (value) of `bnp3` (31).

31.
```
bnp(Agr,[N,R]) --> bs(rel,N,R),bnp3(Agr,N).
```

The relative clause is defined as the other clauses (sentences) but is primarily distinguished by lacking an element (the relativized noun phrase). To distinguish it from other clauses it has the word `rel` in the mode slot: `bs(rel,N,R)`. The `N` in the focus slot is a copy of the correlate which is

percolated down into the relative clause to the functional role which is relativized.

The following is a rule for a relative clause where the subject is relativized. The subject noun phrase is therefore missing in the string of categories but N occurs in the functional representation as it is percolated from the correlate, the `N` in `bnp3(Agr,N)`.

32.
```
bs(rel,N,[subj(N),pred([V,Asp,T]),obj(N2),
advl(A1),advl(A2),advl(A3),co(N,Mo,C,F2)]) -->
{permute([badvp(A1), bvt([V,As]), badvp(A2),bnp3(Agr2,N2),
badvp(A3),{bnlex(Agr1,N,F,[])}, % Agr features from corr
rel(Agr1,Agr2,T)],M)}, % special auxiliary in rel clause
{imbefore(bvt([V,As]),rel(Agr1,Agr2,T),M)},
{last(rel(Agr1,Agr2,T),M)}, % relative auxiliary last
surf(M).
```

Note that the noun phrase in the relative clause is a `bnp3` (which may not contain a relative clause). Thus, a relative clause may not contain another relative clause in this grammar. The purpose of this in our system is to avoid recursion problems.

The rule further makes use of a separate set of auxiliaries (syntactically a combination of an indicative auxiliary such as *du* '3sA-3sE-AUX-PRES' and the relativizer itself, which is the suffix *-en*. While later versions of this grammar will have a more advanced morphological component, the combinations of auxiliary and relative marker are not derived morphologically at this stage, but listed in the lexicon as a relative auxiliary (`rel`). Two such relative auxiliaries are listed in (33).

33.
```
rel(agr(sg,_,erg),agr(sg,_,abs),pres) --> [duen].
rel(agr(sg,_,erg),agr(sg,_,abs),past) --> [zuen].
```

The word order constraint requires the main verb to occur before the relative auxiliary and the relative auxiliary to occur last.

The following (34) illustrates the generation of an ergative noun phrase which involves subject relativization.

34.
```
bnp(Agr,F,X,[])
Agr = agr(sg, _4789, erg), F = [[subj(m(amaia, prop)),
pred([learn, fut, pres]), obj(m(basquelanguage,_)),
advl([]),  advl([]), advl([]), co(m(amaia, prop), _5415,
_5418, _5421)], m(amaia, prop)],
X = [euskara, ikasiko, duen, 'Amaiak']
```

euskara	ikasiko	duen	Amaiak
Basque-(ABS)	learn-FUT	3sA-3sE-AUX-PRES-REL	Amaia-ERG

'Amaia, who will learn Basque …'

Programming intransitive sentences

Basque has two types of sentence patterns without an object. One is the type headed by what is known in generative circles as unergative verbs, which are semantically intransitive but agentive and pattern grammatically with transitive verbs. In traditional work on Basque, they are often treated as a subgroup of transitive verbs. To prevent overgeneration, such verbs are treated in our model as a separate class, termed *semi-transitives*, which combine transitive morphology with an intransitive sentence pattern (35).

35. Amaia-k gaur kantatu du.
    Amaia-ERG today sing have-3sA-3sE
    'Amaia has sung today.'

The verbs in these sentences have been denoted by `bvs`. The auxiliary takes double agreement although there is no overt object. In the rule below we have denoted the object by the empty set `[]` in the functional representation. One may alternatively have no object in the functional representation which would make it more similar to the functional representations of intransitive sentences. There is no object NP in the surface sequence of categories. The agreement value of the auxiliary is set at `agr(sg,_,abs)`, which represents a default value. Note that this sentence type uses a transitive auxiliary.

36.
```
bs(d,Fo,[subj(N1),pred([V,As,T]),obj([]),advl(A1),
advl(A2),advl(A3),co(N2,Mo,C,F2)]) -->
{permute([badvp(A1),bnp(Agr1,N1),bvs([V,As]),
bauxt(Agr1,agr(sg,_,abs),T), badvp(A2),
badvp(A3),btail(C,F2)],M)},
{imbefore(bvs([V,As]),bauxt(Agr1,agr(sg,_,abs),T),M)},
{imbefore(Fo,bvs([V,As]),M)},
{Fo≠badvp([])},
{last(btail(C,F2),M)},
surf(M).
```

Some verb forms pertaining to the class of `bvs` are listed in (37).

37.
```
bvs([dance,perf]) -->[dantzatu].
bvs([dance,imp]) --> [dantzatzen].
bvs([dance,fut]) --> [dantzatuko].
```

An example of a clause generated by the above rule is given below (38).

38.
```
bs(M, Fo, F, X, [])
No.1 : M = d, Fo = bnp(agr(sg, _57823, erg), peru),
F = [subj(peru), pred([dance, perf, pres]), obj([]),
advl([]), advl([]), advl([]), co(_76644,
_64297, [.], [])],
X = ['Peruk', dantzatu, du]
```

Peru-k dantzatu du.
Peru-ERG danced 3sA-3sE-AUX
'Peru has danced.'

Unaccusative sentences

The second type of intransitive verb is that which is referred to in generative work as unaccusative. This is the type which is termed intransitive in traditional work on Basque. The main verb has been denoted by `bvi` in the present paper. An unaccusative verb requires the use of the intransitive auxiliary 'to be' rather than 'to have'. This intransitive auxiliary is denoted by `bauxi`. The sole argument of such a clause is a `bnp` which carries the absolutive case feature and the auxiliary agrees with this. There is no equivalent to a functional object but this role has been included and set at `[]`.

39.
```
bs(d,Fo,[subj(N),pred([V,As,T]),obj([]),advl(A1),
advl(A2),advl(A3),co(N,Mo,C,F2)]) -->
{permute([badvp(A1),bnp(Agr1,N),
bvi([V,As]),bauxi(Agr1,T),badvp(A2),
badvp(A3),btail(C,F2)],M)},
{imbefore(bvi([V,As]),bauxi(Agr1,T),M)},
{imbefore(Fo,bvi([V,As]),M)},
{Fo≠badvp([])}, {last(btail(C,F2),M)}, surf(M).
```

Below we list some `bvi` lexical entries, as well as various forms of the intransitive auxiliary `bauxi` (40).

40.
```
bvi([go,perf]) --> [joan].
bvi([go,imp]) --> [joaten].
bvi([go,fut]) --> [joango].
bvi([come,perf]) --> [etorri].
bvi([come,imp]) --> [etortzen].
bvi([come,fut]) --> [etorriko].
bauxi(agr(sg,_,abs),pres) --> [da].
bauxi(agr(pl,_,abs),pres) --> [dira].
bauxi(agr(sg,_,abs),past) --> [zen].
bauxi(agr(pl,_,abs),past) --> [ziren].
```

Synthetic inflection

Certain Basque verbs can be inflected directly, without using an auxiliary. In order to demonstrate synthetic inflection we include the following rule (41). Note that no auxiliary is included in the pattern, and that the main verb is denoted by `bfvt` ('finite transitive verb').

41.
```
bs(d,Fo,[subj(N1),pred([V,As,T]),obj(N2),advl(A1),
advl(A2),advl(A3),co(N1,Mo,C,F2)]) -->
{permute([badvp(A1),bnp(Agr1,N1), badvp(A2), bnp(Agr2,N2),
bfvt(Agr1,Agr2,[V,As,T]),badvp(A3),btail(C,F2)],M)},
{imbefore(Fo,bfvt(Agr1,Agr2,[V,As,T]),M)},
{Fo≠badvp([])}, {last(btail(C,F2),M)},surf(M).
```

The rule can for instance analyse the example *Peruk Amaia dakus* (‘Peru sees Amaia’) as in (42).

42.
```
bs(M, Fo, F, ['Peruk', 'Amaia', dakus], [])
No.1 : M = d, Fo = amaia, F = [subj(peru),
pred([see, imp, pres]), obj(amaia), advl([]),
advl([]), advl([]), co(peru, _36798, [])], [.], [])]
```

Peruk Amaia dakus
Peru-ERG Amaia-(ABS) see-PRES-3sA-3sE
‘Peru sees Amaia.’

Such rules require the listing of the inflected main verb in the lexicon. The following (43) illustrates some lexical items with `bfvt` verbs.

43.
```
bfvt(agr(sg,_,erg),agr(sg,_,abs),[see,imp,pres]) -->[dakus].
bfvt(agr(pl,_,erg),agr(sg,_,abs),[see,imp,pres]) -->
[dakuste].
```

Ditransitive sentences

Ditransitive verbs such as *eman* ‘give’ take, in addition to the agent in ERG, two object arguments, one in ABS (the object given) and one in DAT (the recipient). The auxiliary agrees in person and number with all three arguments of the clause. This is exemplified below.

44. Peruk liburua Amaiari eman dio.
Peru-ERG book-(ABS) Amaia-DAT give 3sA-3sE-3sD-AUX
‘Peru has given the book to Amaia.’

This type of clause requires a special syntactic pattern, including a ditransitive main verb (`bvd`) and a ditransitive auxiliary (`bauxd`). The following demonstrates the rule for such a syntactic pattern (45).

45.
```
bs(d,Fo,[subj(N1),pred([V,As,T]),obj(N2),dobj(N3),
advl(A1),advl(A2),advl(A3), co(N1,Mo,C,F2)]) -->
{permute([badvp(A1),bnp(Agr1,N1), badvp(A2),
bnp(Agr3,N3), bnp(Agr2,N2),bvd([V,As]),
bauxd(Agr1,Agr2,Agr3,T),badvp(A3), btail(C,F2)],M)},
{imbefore(bvd([V,As]),bauxd(Agr1,Agr2,Agr3,T),M)},
{imbefore(Fo,bvd([V,As]),M)}, % focus before verb
{Fo≠badvp([])}, {last(btail(C,F2),M)}, surf(M).
```

The verbs and auxiliaries required by this pattern must be listed in the lexicon. Some entries are listed in (46).

46.
```
bvd([give,perf]) --> [eman].
bvd([give,imp]) --> [ematen].
bvd([give,fut]) --> [emango].
bauxd(agr(sg,_,erg),agr(sg,_,abs),agr(sg,_,dat),pres) -->
[dio].
bauxd(agr(pl,_,erg),agr(sg,_,abs),agr(sg,_,dat),pres) -->
[diote].
bauxd(agr(sg,_,erg),agr(pl,_,abs),agr(sg,_,dat),pres) -->
[dizkio].
bauxd(agr(pl,_,erg),agr(pl,_,abs),agr(sg,_,dat),pres) -->
[dizkiote].
```

The grammar is also designed to cope with intransitive verbs which take dative objects, as in the following example. However, for reasons of space, we shall not discuss these patterns in the present paper.

47. Liburua Peruri gustatzen zaio.
book-(ABS) Peru-DAT please-IPF 3sA-3sD-AUX
'Peru likes the book.' (lit. 'The book pleases Peru.')

Some morphological rules deriving case forms

A characteristic fact of Basque is that the case endings are suffixed to whole noun phrases including possible articles and attributives. The typical ergative ending *-k* is thus not only suffixed to the name *Peru* to give *Peruk,* but also to *gizona* (the definite form of *gizon* 'man') to give *gizonak* [man-DET-ERG] and to *gizon bat* 'a man' to give the ergative form *gizon batek*. The ergative form of *gizon handi bat,* 'a big man' is *gizon handi batek.*

Following the hierarchy of noun phrases presented earlier it is convenient to add case suffixes to `bnp3` deriving other forms of `bnp3`, to ensure that the case suffix is added after the definite article. Given that the absolutive case form is morphologically unmarked, it is natural to take this as the basis as shown in the following rule deriving ergative forms for singular nouns. The rule states (using the operator :-) that the existence of `bnp3(agr(sg,_,abs), B,[F],[])` implies the existence of `bnp3(agr(sg,_,erg), B,[F1|X],X)` provided that *-k* is added to the absolutive form `F` unless the latter ends in a consonant, in which case *-ek* is added (48).

48.
```
bnp3(agr(sg,_,erg),B,[F1|X],X) :-
bnp3(agr(sg,_,abs),B,[F],[]),
(ccat(W,C,F),cons(C),ccat(F,ek,F1); % ek after consonant
ccat(F,k,F1)). % else add k
```

This rule derives *Peruk* from *Peru*. The form *Peru* is listed in the lexicon as an absolutive form and being a proper name it is also a `bnp3`. The rule can

also derive the ergative definite and indefinite forms, but the situation is more complicated when the ergative ending is to be added to a `bnp3` consisting of several words. This situation can also be handled in Prolog, but we will not present these rules here. The result of such rules would give ergative forms such as e.g. *gizon batek* 'a man-ERG', and *gizon handi batek* 'a big man-ERG'.

Analogously, dative forms can be derived according to the following rule (49), which gives *Amaiari* 'Amaia-DAT' from *Amaia* 'Amaia-ABS', *gizon bati* 'one man-DAT' from *gizon bat* 'one man-ABS'. Here, too, noun phrases including several words would require more complex implication rules.

49.
```
bnp(agr(sg,_,dat),B,[F1|X],X) :-
bnp(agr(sg,_,abs),B,[F],[]),
(ccat(W,C,F),cons(C), ccat(F,i,F1);
% if ending in consonant
ccat(F,ri,F1)). % else add ri
```

The use of these and similar rules can derive the entire case morphology of Basque productively. In the present version, we have only presented rules deriving ergative and dative case.

50.
```
DEMO MT Basque-Chinese (pinyin)

COMMANDS
bs(_, _, F1, [eskutitza,ikasleak,idatzi,du], []),
% from Basque
assert(lex(p)), % Use Pinyin transcription
cs(_, _, F1, X, []). % generate Chinese

RESULTS
F1 = [subj(m(student, sg)), pred([write, perf, pres]),
 obj(m(letter, sg)), advl([]), advl([]),
advl([]), co(m(student, sg), d, [], [])], % interlingua

X = [xue2sheng1, yi3jing1, xie3, xin4, le, .]
% Chinese text
```

Eskutitza ikasleak idatzi du.
letter-(ABS) student-ERG write AUX-3sA-3sE
'The student has written the letter.'

Xuésh2ng y#j3ng xi" xìn le.
student already write letter PRF
'The student has written the letter.'

References

Eeg-Olofsson, Mats & Bengt Sigurd. 2001. 'Permutational Grammar for free word order languages'. *Working Papers* 48, 15-23. Lund University, Dept. of Linguistics.

Holmer, Arthur. 1999. 'An active analysis of Basque ergativity'. *Fontes Linguæ Vasconum* 81, 189-225. Pamplona: Gobierno de Navarra.

Holmer, Arthur & Bengt Sigurd. forthcoming. 'Implementing Basque grammar on the computer'. To appear in: *Fontes Linguæ Vasconum*. Pamplona: Gobierno de Navarra.

Ortiz de Urbina, Jon. 1989. *Parameters in the grammar of Basque – GB approach to Basque syntax*. Dordrecht: Foris.

Sigurd, Bengt (ed.). 1994. *Computerized grammars for analysis and machine translation*. Lund University Press.

Trask, R. L. 1997. *The history of Basque*. London: Routledge.