This text is the pre-final version of a paper published in Léa Nash and Pollet Samvelian (eds.), *Approaches to Complex Predicates* (Bril, 2015, pp. 46-69).

Univerbation of light verb compounds and the Obligatory Coding Principle

The *Obligatory Coding Principle* accounts for the inventories of possible coding frames in languages that, according to the current terminology, can be characterized as consistently accusative or consistently ergative in their system of argument coding. In coding frame inventories fully consistent with the Obligtory Coding principle, every coding frame includes a given type of coding, either A (in *obligatory A coding languages*) or P (in *obligatory P coding languages*). However, languages with coding frame inventories violating this principle are not exceptional. This paper examines the questions raised by light verb constructions with respect to the Obligatory Coding Principle, in particular the possible impact of the univerbation of light verb constructions on argument coding systems initially consistent with the principle or obligatory P coding. The discussion is based on an analysis of the role of the univerbation of light verb compounds in the changes that have affected the situation of Basque with respect to the Obligatory Coding Principle, and a comparison of the situation of Basque with that of Andic languages (East Caucasian).

1. Introduction

This paper examines the possible impact of the univerbation of light verb compounds on argument coding systems. After some terminological clarifications (Section 2), in Section 3, I discuss the typological parameter of Obligatory Coding, a reformulation of the traditional distinction between morphologically accusative and morphologically ergative languages which allows a better understanding of the cross-linguistic variation in argument coding systems. In Section 4, I present the questions raised by light verb constructions with respect to the Obligatory Coding Principle. In Section 5, I discuss the role of the univerbation of light verb compounds in the changes that have affected the situation of Basque with respect to the Obligatory Coding Principle. In Section 6, I compare the situation of Basque with that of Andic languages, a group of languages belonging to the East Caucasian family. In Section 7, I discuss a possible explanation of the contrast observed between Basque and Andic languages. Section 8 summarizes the main conclusions.

2. Some terminological clarifications

2.1. Canonical vs. non-canonical coding frames

Each individual language has an inventory of possible coding frames through which verbs express their argument structure. Formal contrasts between arguments may involve three kinds of coding properties: flagging, indexation, and linear order. A coding frame is considered non-canonical if it is found with a restricted set of verbs. The distinction between canonical and non-canonical coding frames is gradient rather than categorical, and may evolve in the history of a language.

2.2. Transitivity

In the terminology used in this paper, the characterization of verbs as *transitive* and *intransitive* does not refer to the number of essential participants in the events they denote. Verbs encoding events involving one, two, or three essential participants are designated here as *monovalent*, *bivalent*, and *trivalent* verbs, respectively. In the use of the term *transitive* adopted in this paper, the delimitation of the set of transitive verbs is language-specific and relies on formal criteria, but the sets of transitive verbs of the individual languages are universally defined as including a particular semantic class of verbs, the *core transitive verbs*, defined as bivalent verbs expressing meanings compatible with a maximum degree of semantic transitivity.

The coding frame selected by core transitive verbs in uses involving a maximum degree of semantic transitivity is designated as *transitive type of argument coding*, abbreviated as *transitive coding*. The notions of core transitive verb and transitive type of argument coding are comparative concepts in the sense of Haspelmath (2010).

The transitive type of argument coding, identified as such by reference to core transitive verbs, may also be selected by verbs that are not core transitive verbs. The term of *transitive verb* without further specification refers to verbs whose coding frame includes two terms coded like the two arguments of core transitive verbs, whatever the semantic roles they are assigned.

For example, English *see* is not a core transitive verb, but its coding frame identifies it as transitive. Basque *ikusi* 'see' must also be identified as a transitive verb, since its coding frame <ERG, Ø> is the same as that of a core transitive verb such as *puskatu* 'break' – Ex. (1). By contrast, Akhvakh *harigurɯa* 'see' is not transitive, since its coding frame <DAT, Ø> is different from the coding frame <ERG, Ø> selected in Akhvakh by core transitive verbs such as *biq'ōrɯa* 'break' – Ex. (2).[1]

(1) Basque[2]

a. *Haurr-ek ispilu-a puskatu dute.*
child-PL.ERG mirror-SG break.CPLV PRES.3SG.3PL
'The children have broken the mirror.'

b. *Haurr-ek ispilu-a ikusi dute.*
child-PL.ERG mirror-SG see.CPLV PRES.3SG.3PL
'The children have seen the mirror.'

(2) Akhvakh[3]

a. *Mik'i-de istaka biq'^{w}āri.*
child-ERG glass break.CPLV
'The child broke the glass.'

[1] The use of Ø in the schematization of coding frames is explained in Section 2.3.
[2] The Basque examples quoted in this paper have been checked by Céline Mounole.
[3] The Akhvakh examples quoted in this paper come from the author's field notes.

b. *Mik'i-ɬa istaka harig^{w}ari.*
child-DAT glass see.CPLV
'The child saw the glass.'

2.3. Zero case

In languages in which nouns are inflected for case, I designate as *zero case* (represented as Ø in the schematization of coding frames) the case form of nouns that coincides with the form used in isolation for quotation and labeling, whatever the distribution of this form in syntactic contexts. The notion of labeling includes for example the case of nouns accompanying a picture representing a possible referent, nouns written on signal boards, nouns describing the content of a box on which they are written, etc. In most languages, the Zero case is characterized by the absence of an overt case marker, but there are exceptions.[4]

The term of zero case can be understood as a cover term for the case forms currently labeled nominative or absolutive. There are two main reasons for preferring it.[5] On the one hand, the distinction between nominative and absolutive is not really useful, since a nominative case in an unproblematic accusative language is simply a zero case contrasting with an accusative case, and an absolutive case in an unproblematic ergative language is simply a zero case contrasting with an ergative case. On the other hand (and this is crucial), the definition of nominative and absolutive as it is commonly formulated can only lead to inconsistencies if one tries to apply it to less common patterns of case marking.[6]

2.4. Ergative case, ergative alignment, ergative languages

If a case form distinct from the quotation / labeling form of nouns is used to encode the P argument of transitive verbs but not the A argument, this form is commonly labeled *accusative case*, and if a case form distinct from the quotation / labeling form of nouns is used to encode the A argument of transitive verbs but not the P argument, it is commonly labeled *ergative case*.[7] Note that this definition allows using the label 'ergative case', not only for languages in which a marked case form assigned by transitive verbs to their agent is never assigned to arguments of intransitive verbs, but also in the following two types of situation:[8]

---

[4] For example, in Russian, the Zero case (alias Nominative) has no overt ending in the singular for nouns such as *gorod* 'town' but an ending *-a* for nouns such as *mašina* 'car', whose form with no overt ending (*mašin*) expresses the Genitive case in the plural.

[5] See Creissels (2009a) for a more detailed discussion.

[6] In a paper dealing mainly with Basque and Andic languages, a third reason is that, in most descriptions of Andic languages, a case form that meets the definition of 'absolutive' put forward in the recent typological literature is labeled 'nominative', whereas in recent publications on Basque, 'absolutive' refers to a morphological case for which this label is somewhat problematic, since Basque has a substantial class of monovalent verbs that do not assign this case to their sole argument.

[7] *Oblique case* is a possible label for marked case forms involved in the coding of both A and P, depending on factors such as verb inflection. Such case forms can be found in some 'split-ergative' languages, for example Kurmanji Kurdish.

[8] The terms of extended / generalized ergative case marking as I use them here must be understood, in a strictly synchronic perspective, as describing the syntactic distribution of ergative case marking. They do not imply that this distribution results from a historical process of extension of an ergative case

– *extended ergative case marking*, in which a substantial class of verbs whose coding frame includes no P term assign to one of their arguments the same marked case form as that assigned by transitive verbs to their A argument;
– *generalized ergative case marking* (commonly designated as the 'marked-nominative' type of case-marking), in which all coding frames must include a term in the same marked case form as the A argument of transitive verbs.

By contrast, I will avoid using 'accusative' and 'ergative' as labels for types of alignment. The main reason is that the extension of ergative marking to arguments of intransitive verbs results in situations in which ergative case marking is assigned by a class of intransitive verbs commonly characterized as 'unergatives', and that the terminology currently used by typologists leads to characterize as 'aligned accusatively'. Such a terminological mess can only result in misunderstandings and analytical errors. For example, one cannot be satisfied with a terminological system according to which the only possible characterization of the situation of a language like Basque is that the increase in the proportion of intransitive verbs assigning ergative marking triggers the gradual disappearance of ergative alignment. This is the reason why I propose the unambiguous terms of *A-alignment* and *P-alignment* for the types of alignment currently designated in the typological literature as accusative alignment and ergative alignment, respectively.

For similar reasons, I reject the use of accusative and ergative as possible labels for a global characterization of argument coding systems. According to the definitions found in the recent literature, a morphologically ergative language should be identified as such with reference to its alignment properties only, but in current practice, it is clear that for many linguists, the notion of ergative language refers to a bundle of features:

(a) FLAGGED AGENTS, i.e. the coding of the agents of core transitive verbs by means of either an adposition or a case form (commonly termed *ergative case*) distinct from the zero case used in isolation for quotation or labeling;
(b) UNFLAGGED PATIENTS;
(c) EITHER NO INDEXATION AT ALL, OR INDEXATION OF PATIENTS ONLY;
(d) OBLIGATORY P CODING, i.e. the selection of P coding as the default type of argument coding that must be included in the coding frame of all verbs (and is consequently the only possible coding of sole arguments of monovalent verbs).

These four features tend to co-occur cross-linguistically, but are nevertheless logically independent, and are dissociated in some languages, which leads to inconsistencies in

---

whose use was initially limited to the A argument of transitive verbs. There is clear evidence that such a historical process was responsible for the extended ergative case marking found in Basque (as will be discussed further in this paper), or in Kartvelian languages, but other scenarios can be imagined, and for the languages of East Africa whose argument coding systems involve generalized ergative case marking, the extension of the use of an ergative case initially limited to the A argument of transitive verbs is certainly not the most plausible historical explanation.

the characterization of languages in which such a dissociation occurs.[9] The solution I propose is to replace ergative and accusative as labels characterizing systems of argument coding by the unambiguous terms of *A-unmarked* vs. *P-unmarked* systems of argument coding, conceived as referring to prototypes conflating features that tend to co-occur cross-linguistically but can nevertheless be dissociated in individual languages.

3. The Obligatory Coding Principle

Accusativity / ergativity is commonly defined in terms of alignment of the sole argument of monovalent verbs with either the agent or the patient of core transitive verbs, but the alignment of the coding characteristics of S with those of either A or P can be viewed as a particular case of a more general principle, the Obligatory Coding Principle, according to which every coding frame in a given language must include, either (in 'accusative' languages) a term with coding properties identical to those of the A argument of transitive verbs, or (in 'ergative' languages) a term with coding properties identical to those of the P argument of transitive verbs. In this paper, these two situations will be characterized as *obligatory A coding* and *obligatory P coding*, respectively.

However, many languages have inventories of possible coding frames hardly compatible with the Obligatory Coding Principle. For example, Basque has two subsets of monovalent verbs, some of them assigning P coding to their sole argument – Ex. (3c), and the others assigning A coding – Ex. (3b).

(3) Basque

a. *Haurr-ak* *ur-a* *ekarri* *du.*
child-SG.ERG water-SG bring.CPLV PRS.3SG.3SG
'The child brought the water.'

b. *Ur-ak* *irakin* *du.*
water-SG.ERG boil.CPLV PRS.3SG.3SG[10]
'The water boiled.'

c. *Haurr-a* *etorri* *da.*
child-SG come.CPLV PRS.3SG
'The child came.'

A formal elaboration of the Obligatory Coding Principle is found in the generative literature under the name of Obligatory Case Parameter (Bobaljik 1993, Laka 1993 &

[9] For example, in recent publications on Basque, this language is often characterized as an 'ergative language of the active type'. This formulation is nothing else than a pure and simple *contradictio in terminis*, if ergative and active are taken with their current definitions. It can only be consistent with an understanding of 'ergative language' according to which the overt flagging of agents is more important than the alignment properties of intransitive verbs.

[10] *Du* is a form of the so-called transitive auxiliary, which in principle expresses agreement with two arguments. Verbs that have a sole argument in the Ergative case are conjugated by means of this auxiliary, and agreement with a term in the Zero case takes the default value 3SG.

2000, Rezac 2008a & 2008b). The idea is that the difference between obligatory A coding and obligatory P coding depends on whether a high head, T°, or a lower head, v°, is active for obligatory case assignment (T° active → obligatory A coding, v° active → obligatory P coding). A question that has been particularly discussed, mainly with reference to Basque, is the explanation of the violations of the Obligatory Case Parameter in 'ergative' languages. The solutions that have been proposed draw on the insight that some superficially intransitive verbs may be underlyingly transitive.

I will not discuss this issue further, since this paper is not devoted to an elaboration of the formal aspects of the question, but to an examination of diachronic processes likely to be responsible for the emergence of violations of the Obligatory Coding Principle.

4. Light verb compounds and the Obligatory Coding Principle

Some languages have a particularly high proportion of predicates expressed as light verb compounds in which the light verb is a transitive verb (most often a verb with the meaning 'do, make', as in Ex. (4)), and the non-verbal element is a noun encoded like the P argument of transitive verbs (Samvelian 2012: 16).

(4) Basque

a. *Haurr-ek* *lo* *egiten* *dute.*
child-PL.ERG sleep do.ICPLV PRS.3SG.3PL
'The children are sleeping (lit. are doing sleep).'

b. *Gizon* *horr-ek* *ez* *du* *euskar-az* *hitz* *egiten.*
man DEM.SG-ERG NEG PRS.3SG.3SG Basque-SG.INSTR word do.ICPLV
'This man does not speak Basque (lit. does not do word in Basque).'

The coding frame of such predicates can be schematized as A (X) p V, where V is the verbal element of the light verb compound, (lower-case) p represents the non-verbal element of the compound, coded as if it were the P argument of a transitive verb, A represents an argument to which A coding is assigned, and (X) represents possible additional terms whose presence depends on the argument structure of the predicate, and to which an oblique-like coding is assigned.

In languages with obligatory A coding, considering the non-verbal element of the compound as a term in the construction of the light verb or considering the compound p V as a whole as the syntactic equivalent of a simplex verb does not change anything with respect to the Obligatory Coding Principle, since the construction includes a participant encoded like the A argument of a transitive verb. By contrast, in languages with obligatory P coding, the principle is formally satisfied by the nominal element of the compound, insofar as it is considered a term in the construction of the light verb, but the principle is violated if the complex predicate is taken as a whole, and only NPs representing participants are considered terms of the predicative construction.

Diachronically, there is a general tendency toward fusion of the two elements of such compounds. This univerbation process converts formally transitive constructions schematizable as A (X) p V into constructions schematizable as A (X) V, with a term

showing A coding but no term showing P coding. In languages with obligatory A coding, the constructions resulting from this evolution are perfectly canonical constructions (since they include a term showing A coding), whereas in languages with obligatory P coding, the same process results in a violation of the Obligatory Coding Principle (since the outcome of the evolution is a construction in which no term shows P coding). Interestingly, as will be discussed in the following sections, some languages with obligatory P coding show a strong tendency toward regularization of the non-canonical coding frames resulting from this process, whereas others tend to keep them unchanged.

5. The Basque conspiracy against obligatory P coding

5.1. Introductory remarks

In Basque, transitive verbs uniformly assign the Ergative case to their A argument. As regards the encoding of the P argument of transitive verbs, some varieties (including Standard Basque) uniformly use the Zero case, whereas others have developed a DOM system characterized by an alternation between the Zero case and the Dative case.

Old Basque as attested by texts from the 15th and 16th centuries was a language obeying the principle of obligatory P coding in a relatively strict way. Verbs used in coding frames with no term in the Zero case (in particular, monovalent verbs assigning the Ergative case to their sole argument, such as *irakin* 'boil') were not totally lacking, but they represented a minute percentage of the verbal lexicon – Mounole (2011), and it is reasonable to assume that their construction resulted from isolated accidents in the evolution of individual verbs. However, isolated accidents cannot explain the important increase in the proportion of verbs with coding frames with no term in the Zero case that occurred in the history of most Basque varieties.[11]

As already illustrated by Ex. (4) above, Basque makes wide use of light verb compounds consisting of a bare noun and the verb *egin* 'do, make'. Taken as a whole, such compounds are predicates whose argument structure does not involve an argument encoded as a NP in the Zero case. The argument structure of light verb compounds like *lo egin* 'sleep' or *hitz egin* 'speak' can be represented as <ERG, ø>, where (uppercase) ERG symbolizes the slot for the argument of the light verb compound taken as a whole, and (lowercase) ø symbolizes the slot for the non-verbal element of the compound, whose coding characteristics are similar to those of the P argument of transitive verbs. As observed by Etxepare (2003: 397), such compounds "are not instances of incorporation ... the bare nominal and the verb *egin* can be separated by a number of syntactic operations, and the bare nominal can take partitive case" – see Oyharçabal (2007) for a more detailed analysis of Basque light verb compounds.

[11] For a more detailed discussion of the questions addressed in the rest of this Section, in particular on the role of borrowing in the extension of Ergative marking in Basque, see Creissels & Mounole (2012).

5.2. Simplex verbs whose root coincides with the non-verbal element of a light verb compound

Basque has a few verbs more or less recognizable as originating from the coalescence of the sequence constituted by the two elements of a light verb compound, and some of them at least have coding frames implying that the construction resulting from the incorporation of the bare noun was subsequently regularized. For example, *atzeman* 'seize, get', with the coding frame <ERG, Ø> probably resulted from the univerbation of a light verb construction *(h)atz eman* with a coding frame <ERG, ALL, ø> or <ERG, DAT, ø>, lit. 'put finger (on)'.[12] However, most light verb compounds correspond to simplex verbs whose root simply coincides with the non-verbal element of the compound. In such cases, 'univerbation' does not involve the coalescence of the two elements of a light verb compound, but rather the conversion of the non-verbal element into a verb stem. *Bultza egin* lit. 'do impulse' / *bultzatu* 'push' – Ex. (5) – illustrates this kind of relationship between light verb compounds and simplex verbs. [13]

(5) Basque[14]

a. *Mutil-ak* *ate-ari* *bultza* *egin* *zion.* <ERG, DAT, ø>
boy-SG.ERG door-SG.DAT impulse do.CPLV PST.3SG.3SG.3SG
'The boy pushed the door.'

b. *Mutil-ak* *ate-a* *bultzatu* *zuen.* <ERG, Ø>
boy-SG.ERG door-SG push.CPLV PST.3SG.3SG
same meaning as (a)

In this example, a light verb compound selecting the coding frame <ERG, DAT, ø> corresponds to a simplex transitive verb, which means that the Dative argument of the light verb compound represents the same participant as the argument of the simplex verb in the Zero case. However, in most cases, the arguments of the simplex verb are encoded in the same way as in the light verb construction. Consequently, verbs cognate with the non-verbal element of *egin*-compounds constitute an important proportion of the verbs whose coding frame does not include a term in the Zero case. For example, the light verb compound *dirdir egin* 'shine' and the corresponding simplex verb *dirdiratu* equally assign the Ergative case to their argument – Ex. (6).

---

[12] In present-day Basque, *eman* is the translational equivalent of English 'give', but there is evidence that its original meaning was something like 'put'.

[13] *Bultzatu* is the completive participle, used in Basque grammars and dictionaries as the quotation form of verbs. It can be decomposed as *bultza-* (root) plus *-tu* (completive aspect marker).

[14] The form taken by the auxiliary in this example calls for the following observation. *Zuen* is the past form of the transitive auxiliary (i.e., the auxiliary indexing a term in the Zero case and a term in the Ergative case) which corresponds to the present form *du* found in other examples. *Zion* is the corresponding form of the transitive-with-dative auxiliary, which indexes a third term in the Dative case.

(6) Basque

a. *Eguzki-ak* *dirdir* *egiten* *du.* <ERG, ø>
sun-SG.ERG shining do.ICPLV PRS.3SG.3SG
'The sun is shining.'

b. *Eguzki-ak* *dirdiratzen* *du.* <ERG>
sun-SG.ERG shine.ICPLV PRS.3SG.3SG
same meaning as (a)

The variation observed in the coding frames of simplex verbs cognate with the non-verbal element of an *egin*-compound can be viewed as the result of the interaction between two conflicting tendencies: a tendency to align the encoding of the arguments of the simplex verbs with the encoding of the same arguments in the light verb construction, as in Ex. (6), and a tendency to organize the coding frame of the simplex verb according to the principle of obligatory P coding, as in Ex. (5).[15]

5.3. Borrowings

In present-day Basque varieties, most of the verbs occurring in coding frames with no term in the Zero case are, either verbs cognate with the non-verbal element of a light verb compound, or borrowings. The contribution of light verb constructions to the increase in the proportion of verbs occurring in coding frames with no term in the Zero case results from the tendency to encode the arguments of the simplex verbs corresponding to a light verb construction in the same way as in the light verb construction. As regards borrowing, its contribution to the increase in the proportion of verbs selecting coding frames in contradiction with the rule of obligatory P coding results from a very strong tendency to assign Ergative coding to the argument of all monovalent verbs borrowed from Spanish or French that do not correspond to so-called pronominal verbs (i.e. verbs to which the clitic *se* is attached) in Spanish or French, and to reserve Zero coding for the argument of borrowed monovalent verbs that correspond to pronominal verbs in Spanish or French.

This rule probably originates in the functional equivalence between the absence of the clitic *se* in Romance languages and the use of Ergative coding in Basque in the construction of verbs involved in the causative/anticausative alternation, like *puskatu* 'break': for such verbs, the use of the so-called transitive auxiliary in Basque (implying indexation of two arguments and Ergative coding for one of the arguments) and the absence of *se* in Spanish of French equally mark that the argument structure includes an agent, whereas the use of the so-called intransitive auxiliary in Basque

[15] The first tendency is illustrated above by a semantically bivalent predicate, and the second one by a monovalent predicate, but this plays no role in the resolution of the conflict between these two tendencies. Coding frames lacking a term in the Zero case may equally be maintained or 'regularized' regardless of the semantic valency of the predicate. The only difference is that, if the light verb construction has the coding frame <ERG, DAT, ø> (as in Ex. (5)), Ergative marking is maintained, and the regularization process consists in substituting Zero marking for Dative marking, whereas with simplex verbs cognate with light verb compounds having the coding frame <ERG, ø>, the only possible regularization consists in substituting Zero marking for Ergative marking.

(implying indexation of one argument only, and incompatible with Ergative coding) and the use of *se* in Spanish or French equally mark that the agent is suppressed from argument structure, as illustrated by Ex. (7).

(7) Basque / French

a. *Ispilu-a puskatu dute. / Ils ont cassé le miroir.*
mirror-SG break.CPLV PRS.3SG.3PL they have broken the mirror
'They broke the mirror.'

b. *Ispilu-a puskatu da. / Le miroir s'est cassé.*
mirror-SG break.CPLV PRS.3SG the mirror has_itself broken
'The mirror broke.'

The awareness of this equivalence is certainly the reason why Basque speakers borrowing monovalent verbs from French or Spanish tend to model the choice between Zero and Ergative coding of the argument on the distinction between pronominal and non-pronominal verbs in French or Spanish, which results in a considerable increase in the number of monovalent verbs with Ergative coding of their sole argument.

5.4. From (relatively) strict to (relatively) loose ergative coding: the verbs of aiming

Borrowing and the creation of simplex verbs cognate with light verb compounds are not sufficient to explain the changes that affected the implementation of the Obligatory Coding Principle in Basque, and must rather be viewed as elements of a more general 'conspiracy' toward development of types of coding frames which in Old Basque had a rather marginal status. Crucial evidence is provided by a semantically consistent group of verbs that were neither borrowed nor created from light verb constructions during the attested history of Basque, and whose coding frame has undergone evolutions that reveal a deep change in the principles underlying the use of Ergative coding: the verbs of aiming (i.e., the verbs referring to events involving a participant exerting a volitional activity directed toward another participant, without however triggering a change of state affecting the second participant: *help*, *follow*, *beg*, *attack*, etc.).

In the oldest Basque texts, aiming verbs are typically found with the coding frame <Ø, DAT>, with the aimer in the Zero case, but no modern dialect has maintained this situation. In all dialects, the aimer tends to show Ergative coding, but variation can be observed in the treatment of the second participant: Western dialects have maintained the ancient Dative coding of the second argument, resulting in a pattern <ERG, DAT> that violates the principle of obligatory P coding – Ex. (8a), whereas in Eastern dialects, the coding of both terms has changed, and the original pattern <Ø, DAT> has been replaced by the canonical pattern <ERG, Ø> – Ex. (8b).

(8) Basque (a: Western varieties, b: Eastern varieties)

a. *(Ni-k) Amaia-ri bazkari-a prestatzen lagundu nion.*
1SG-ERG Amaia-DAT lunch-SG prepare.ICPLV help.CPLV PST.3SG.3SG.1SG
'I helped Amaia prepare the lunch.'

b. *(Ni-k) Amaia bazkari-a prestatzen lagundu nuen.*
1SG-ERG Amaia lunch-SG prepare.ICPLV help.CPLV PST.3SG.1SG
'I helped Amaia prepare the lunch.'

The use of <Ø, DAT> as the coding frame for aiming verbs in Old Basque suggests that, at some point in its history, the ancestor of Basque was probably a language characterized not only by obligatory P coding, but also by 'strict' ergative coding, according to a distinction between strict and loose ergative coding introduced by Harris (1985) and applied to Basque by Aldai (2008).

In languages with strict ergative coding, ergative coding tends to be limited to agents of core transitive verbs in contexts implying a high degree of semantic transitivity, whereas in languages with loose ergative coding, ergative coding is widely used to encode the most agent-like argument of bivalent verbs that are not core transitive verbs, irrespective of the precise semantic roles they assign and of the contexts in which they are used.

The Zero coding of aimers found in Old Basque is characteristic of systems close to the prototype of strict ergative coding, since aimers differ from typical agents in that their activity does not result in a change of state affecting the participant toward which it is directed. However, Old Basque also showed features typical of systems with loose ergative coding, in particular Ergative encoding of experiencers of perception verbs. For example, *ikusi* 'see' (illustrated by Ex. (1) above) is already found in the oldest Basque texts with its experiencer in the same Ergative case as in present-day Basque. This suggests that the tendency to extend Ergative coding already existed.

To conclude this section, the change in the coding frames of aiming verbs observed in historical Basque can be viewed as the elimination of one of the last vestiges of a more ancient system characterized by strict ergative coding. In comparison with other languages whose noun inflection includes an ergative case, the Ergative case of Basque is not very marked semantically, and this paves the way for the development of coding frames in which an Ergative term does not necessarily contrast with a term in the Zero case, in particular via the creation of simplex verbs cognate with the non-verbal element of light verb compounds.

6. The case of Andic languages

Contrary to Basque, Andic languages (a group of closely related Nakh-Daghestanian languages spoken in the western part of Daghestan) obey the principle of obligatory P coding in a relatively strict way. For example, among the 8000 headwords of the Akhvakh-Russian dictionary (Magomedova & Abdulaeva (2007)), I have found only 15 verbs with coding frames including no term in the Zero case. Almost all of them are verbs of aiming, and it is interesting to observe that, in languages with obligatory

P coding, this semantic type of verbs shows a particularly marked tendency to violate the rule according to which every coding frame must include a P term. The case of Basque has already been evoked in Section 5, and in addition to East Caucasian languages, the same phenomenon can be observed for example in Kurmanji Kurdish. A plausible explanation is that the coding frames that most faithfully reflect the argument structure of aiming verbs are those in which the aimer is encoded like typical agents, and the aimee like the argument of movement verbs expressing destination of movement. In languages with obligatory A coding, such coding frames are perfectly canonical, whereas in languages with obligatory P coding, they violate the Obligatory Coding Principle.

Not only in Akhvakh, but more generally in Andic languages, most exceptions to the rule of obligatory P coding are aiming verbs that assign the Ergative case to the aimer and the Allative or Locative case to the aimee: ‘look at’ – Ex. (9), ‘listen’ – Ex. (10), ‘bite’ – Ex (11), ‘pinch’ – Ex. (12), ‘sting’ – Ex. (13).

(9) Akhvakh
Wašo-de di-ga eqāri.
boy-ERG 1SG-ALL look_at.CPLV
‘The boy looked at me.’

(10) Tindi (Magomedova 2003)
Di-qa anix̄^{j}ā hik’^{j}i oš̌^{w}-ī
1SG-LOC/ALL listen.ICPLV NEG DEM.M-ERG
‘He does not listen to me.’

(11) Karata (Magomedova & Xalidova 2001)
χ^{w}aj-ol q̄’^{w}are di-č’o.
dog-ERG bite.CPLV 1SG-LOC
‘The dog bit me.’

(12) Tindi (Magomedova 2003)
Oš̌^{w}-i č’uno di-č’i.
DEM.M-ERG pinch.CPLV 1SG-LOC/ALL
‘He pinched me.’

(13) Akhvakh
Di-ge *ī'iž^{w}ali-de* *č'ināri*
1SG-LOC wasp-ERG sting.CPLV
'A wasp stung me.'

The variation observed in the expression of these meanings in Andic languages supports a hypothesis already suggested by Charachidzé (1981) for Avar, according to which these exceptional coding frames may result from the reduction of the canonical frames <ERG, Ø, ALL> or <ERG, Ø, LOC> regularly used for verbs expressing meanings of the type 'X applies/holds Y on Z' (and found in particular with the verbs expressing 'hit', with the hittee in the Locative or Allative case and the instrument in the Zero case).

This hypothesis is supported by the fact that not all Andic languages have simplex bivalent verbs with meanings such as 'look at', 'listen', 'bite', 'sting', 'pinch'. For example, Tindi expresses 'sting' as *eq̄^{w}a k̄^{w}ēt^{j}a*, lit. 'hit the sting (on someone)', and 'bite' as *saldi bix̄^{j}it^{j}a*, lit. 'hold the teeth (on someone)', with respectively *eq̄^{w}a* 'sting (noun)' and *saldi* 'teeth' in the Zero case. In at least two cases, 'listen' and 'bite', there is clear evidence that the simplex verbs found in some Andic languages result from the univerbation of such light verb compounds.

In the case of 'listen', three situations are found among Andic languages:

– Some Andic languages express 'listen' by means of a construction involving the noun 'ear' in addition to the noun phrases encoding the two participants, as in Godoberi *hãt'uk'ja rik̄i*, literally 'fix the ear (on someone/something)' – Ex. (14). Formally, this construction is an instance of the regular coding frame <ERG, Ø, ALL> with *hãt'uk'ja* 'ear' in the Zero case.
– Others have a verb 'hear' with the exceptional coding frame <ERG, ALL>: Tindi *anix̄^{j}it^{j}a* – Ex. (10) above, repeated here as (15), Chamalal *wołuk'la*, Bagvalal *aštila*;
– A verb 'hear' with the regular coding frame <Ø, ALL> is found in two Andic languages: Akhvakh *hãdax̄uruɬa* – Ex. (16), Karata *ãdukała* – Ex. (17).

(14) Godoberi (Saidova 2006)
Wašu-di *imu-q̄i* *hãt'uk'^{j}a* *rikki* *rukkida.*
son-ERG father-LOC/ALL ear hold.INF must.ICPLV
'The son must listen to his father.'

(15) Tindi (Magomedova 2003)
Di-q̄a *anix̄^{j}ā* *hik'^{j}i* *oš̌^{w}-ī*
1SG-LOC/ALL listen.ICPLV NEG DEM.M-ERG
'He does not listen to me.'

(16) Akhvakh
Waša *imo-ga* *hãdax̄ari.*
boy father-ALL listen.CPLV
'The boy listened to his father.'

(17) Karata (Magomedova & Xalidova 2001)
Waša imo-χar ãduke.
boy father-ALL listen.CPLV
'The boy listened to his father.'

Interestingly, 'fix the ear on' is the obvious etymology of Akhvakh *hãdaxuruɬa* (compare with *hãde* 'ear', *bixuruɬa* 'fix' – the root of this verb is *-ix-*), in spite of the fact that the NP representing the listener is assigned the Zero case instead of the Ergative case that should be expected from this etymology.

The variation in the expression of 'listen' in Andic languages provides therefore evidence supporting the reconstruction of the following evolution:

– at a first stage, the coalescence of a trivalent verb selecting the regular frame <ERG, Ø, ALL> with a noun in the Zero case creates a bivalent verb with the exceptional coding frame <ERG, ALL>;
– at a second stage, attested by Akhvakh and Karata, the exceptional coding frame resulting from this evolution may be regularized into <Ø, ALL>.

A similar variation is attested in the expression of 'bite' too, with however a different coding frame in the languages in which 'bite' has a regular coding frame:

– Some Andic languages express 'bite' by means of a construction involving a noun phrase with the meaning 'tooth' in addition to those encoding the two participants, as in Tindi *saldi biχʲiɬʲa*, lit. 'hold the teeth (on someone)'.
– Others have a verb 'bite' with the exceptional coding frame <ERG, LOC>: Karata *q̄'ʷarała* – Ex. (11) above, repeated here as (18), Akhvakh *q̄'eleč'uruɬa* – Ex. (19), Chamalal *q̄'āna* – Ex. (20);
– A verb 'bite' with the regular coding frame <ERG, Ø> is found in two Andic languages: Godoberi *q̄'ami* – Ex. (21), and Bagvalal *salīla* – Ex. (22).

(18) Karata (Magomedova & Xalidova 2001)
χʷaj-ol q̄'ʷare di-č'o.
dog-ERG bite.CPLV 1SG-LOC
'The dog bit me.'

(19) Akhvakh
χʷe-de di-ge q̄'eleč'ari.
dog-ERG 1SG-LOC bite.CPLV
'The dog bit me.'

(20) Chamalal (Magomedova 1999)
χʷāj-d q̄'ānnidaošu-č'.
dog-ERG bite.CPLV DEM.M-LOC
'The dog bit him.'

(21) Godoberi (Saidova 2006)
χʷaji-di q̄'ami how.
dog-ERG bite.CPLV DEM.M
'The dog bit him.'

(22) Bagvalal (Magomedova 2004)
ʕamo-r salli dib lela.
donkey-ERG bite.CPLV 1SG.GEN.N hand
'The donkey bit my hand.'

Interestingly, Bagvalal *sallila* 'bite' is quite obviously cognate with *salʷ* 'tooth', which suggests that this verb results from the univerbation of a compound similar to Tindi *saldi biχʲiɬʲa*, in spite of the fact that the Zero case assigned to the NP representing the bitee does not correspond to what could be expected from this etymology.

The variation in the expression of 'bite' in Andic languages provides evidence supporting the reconstruction of the following evolution:

– at a first stage, the coalescence of a trivalent verb occurring in the regular frame <ERG, Ø, LOC> with a noun in the Zero case creates a bivalent verb with the exceptional coding frame <ERG, LOC>;
– at a second stage, attested by Bagvalal and Godoberi, [16] the exceptional coding frame resulting from this evolution may be regularized into <ERG, Ø>.

The following observations about the verbs expressing 'bite' in Andic languages confirm that some aspects of semantic transitivity may play a role in such evolutions:

– Several Andic languages have verbs with the meaning 'eat' that are reflexes of the root **q̄'am* whose reflexes are glossed 'bite' in the dictionaries of other languages: Akhvakh *q̄'ōnuɬa* 'eat', Karata *q̄'amaɬa* 'eat', Bagvalal *q̄'anila* 'eat'. Interestingly, the irregular coding frame observed with the reflexes glossed 'bite' is not observed with those glossed 'eat', which select the coding frame <ERG, Ø> characteristic of core transitive verbs.
– As indicated above, the Akhvakh verb *q̄'eleč'uruɬa* 'bite' is used in the coding frame <ERG, LOC> to encode prototypical biting events (for example, 'The dog bit me'). However, it may also behave as a transitive verb with the coding frame <ERG, Ø>, when it refers to biting events that affect the physical integrity of the second participant (bite off a piece of something and eat it), as in Ex. (23).

(23) Akhvakh
Wašo-de ʕeče q̄'eleč'ari.
boy-ERG apple bite.CPLV
'The boy bit off a chunk of apple.'

[16] According to Daniel (2001), verbs with coding frames including no term in the Zero case are particularly rare in Bagvalal.

Andic languages make a wide use of light verb compounds consisting of a transitive verb and a noun in the Zero case, but have very few verbs with coding frames violating the obligatory P coding principle, and none of them is used with a non-canonical coding frame in all Andic languages. The obvious conclusion is that Andic languages have a strong tendency to regularize the non-canonical coding frames arising from the univerbation of light verb compounds. Interestingly, exceptions are mostly found among aiming verbs, which independently from the processes examined here show a particular propensity to violate the obligatory P coding principle, as evidenced among others by Basque.

Therefore the question is why, in some of the languages with obligatory P coding that make a wide use of light verb constructions, there is a very strong tendency to eliminate the violations of the Obligatory Coding Principle that arise from the univerbation of light verb compounds whose non-verbal element is a noun encoded as if it represented a patient, whereas in others, the univerbation of light verb compounds contributes to an increase in the proportion of verbs with coding frames that are not compatible with the principle of obligatory P coding.

7. Discussion

I have argued above that, in the history of Basque, the weakening of the tendency toward regularization of coding frames contradicting the principle of obligatory P coding must be viewed as part of a general trend toward relaxation of the constraints limiting the use of Ergative coding. And precisely, in this respect, Andic languages are strikingly different from Basque.

East Caucasian languages in general, and Andic languages in particular, have a strong tendency to use bivalent verbs that are not core transitive verbs with coding frames other than the <ERG, Ø> coding frame that characterizes transitive verbs. For example, in Basque, experiencers in the Ergative case are common, and on this point, the situation of Old Basque was not different. By contrast, the Ergative cases of Andic languages are not used to encode experiencers. As illustrated by Ex. (24), in Andic languages, most verbs of perception, cognition, or emotion assign the Dative case to their experiencer and the Zero case to their other argument.

(24) Akhvakh

a. *Di-ɬa* *miq̄’i* *harig^{w}ari*
1SG-DAT road see.CPLV
‘I saw the road.’

b. *Di-ɬa* *huduwe* *woq’ido.*
1SG-DAT DEM.M know.ICPLV
‘I know him.’

It is true that, in itself, the wide use of light verb constructions already constitutes a deviation from strict ergative coding, since it results in assigning ergative coding to participants that may have very few in common semantically with typical agents. However, the Andic languages show that, in so far as this deviation from the prototype of strict ergative coding remains isolated, it can co-exist with a strong

tendency to regularize the coding frames including no term in the Zero case that arise from the univerbation of light verb compounds. As suggested by Basque, this tendency can only be weakened as part of a more general trend toward 'loose' ergative marking.

8. Conclusion

As already mentioned in Section 2.4, cross-linguistically, there is a strong correlation between four features that, taken together, define a prototype corresponding to what seems to be the most widespread understanding of the term 'ergative language':

(a) FLAGGED AGENTS, i.e. the coding of the agents of core transitive verbs by means of either an adposition or a case form (commonly termed *ergative case*) distinct from the zero case used in isolation for quotation or labeling;
(b) UNFLAGGED PATIENTS;
(c) EITHER NO INDEXATION AT ALL, OR INDEXATION OF PATIENTS ONLY;
(d) OBLIGATORY P CODING, i.e. the selection of P coding as the default type of argument coding that must be included in the coding frame of all verbs (and is consequently the only possible coding of sole arguments of monovalent verbs).

The correlation is however not absolute, hence the terminological problems raised by the indiscriminate use of 'ergative' with reference to a type of transitive coding and a type of alignment. In order to avoid the misunderstandings resulting from this terminological practice, I have proposed to characterize the type of alignment commonly termed ergative alignment as *P-alignment* (as opposed to *A-alignment*), and similarly, the prototype conflating the four features enumerated above can be characterized as *P-unmarked system of argument coding* (as opposed to *A-unmarked system of argument coding*).

Not all languages have systems of argument coding lending themselves to a straightforward classification as P-unmarked or A-unmarked systems. In some languages, the coding of both agents and patients involves the use of marked case forms or adpositions. The use of a marked case form for agents is also found in some split-S languages, such as Georgian or Basque, in which a sizeable proportion of monovalent verbs assign to their sole argument the same marked case as that assigned by transitive verbs to their agent. The use of a marked case form is also found in flagged-A/S languages (more commonly known as 'marked-nominative' languages), in which a marked case form used for the agent of transitive verbs is also used to encode the sole argument of all monovalent verbs, whereas an unmarked case form is used for patients.[17]

In this paper, I have tried to analyze the role of the univerbation of light verb compounds in the evolutions that may affect the argument coding system of languages initially close to the prototype of P-unmarked system of argument coding, with in particular flagged agents, unflagged patients, and few exceptions to the rule

[17] Flagged-A/S languages, characterizable also as languages with generalized ergative marking (see Section 2.4) are very common in Africa, but extremely rare elsewhere in the world. On the flagged-A/S languages of Africa, see König (2008).

of obligatory P coding. If no regularization of the non-canonical coding frames resulting from the univerbation of light verb compounds occurs, this evolution mechanically results in a shift from the P-unmarked type of argument coding into extended ergative marking (and possibly, at a later stage, into generalized ergative marking as defined in Section 2.4 above). However, the comparison of Basque and Andic data suggests that the creation of simplex verbs cognate with light verb compounds can only contribute to such a shift as part of a more general 'conspiracy'. As evidenced by Andic languages, in languages that stand relatively close to the prototype of strict ergative coding, with an ergative case relatively marked semantically, the coding frames contradicting the obligatory P coding principle that arise as the automatic result of the univerbation of light verb compounds tend rather to change in order to comply with the principle of obligatory P coding.

Abbreviations

A: agent, ACC: accusative, ALL: allative, CPLV: completive aspect, DAT: dative, DEM: demonstrative, ERG: ergative, F: feminine, GEN: genitive, ICPLV: incompletive aspect, INF: infinitive, INSTR: instrumental, LOC: locative, M: masculine, N: neuter, NEG: negation, OBL: oblique stem, Ø: zero case, P: patient, PRS: present, PST: past, S: sole argument of monovalent verbs, SG: singular, V: verb, X: oblique

References

Aldai, Gontzal. 2008. 'From ergative case marking to semantic case marking: the case of historical Basque'. In Mark Donohue and Søren Wichmann (eds.) *The typology of semantic alignment*. Oxford: Oxford University Press. 197-218.

Bobaljik, Jonathan David. 1993. 'On ergativity and ergative unergatives'. *MIT Working Papers in Linguistics* 19. 45-88.

Charachidzé, G. 1981. *Grammaire de la langue avar*. Paris: Editions Jean-Favard.

Creissels, Denis. 2008. 'Direct and indirect explanations of typological regularities: the case of alignment variations'. *Folia Linguistica* 42-1. 1-38.

Creissels, Denis. 2009a. 'Uncommon patterns of core term marking and case terminology'. *Lingua* 119. 445-459.

Creissels, D. 2009b. 'Valency properties of Northern Akhvakh verbs.' http://www.deniscreissels.fr/public/Creissels-valency_classes_project_Akhvakh.pdf

Creissels, Denis & Céline Mounole. '2012. Non-canonical valency patterns in Basque, variation and evolution'. Paper presented at the 45th Annual Meeting of the Societas Linguistica Europaea (Stockholm, 29 August – 1 September 2012). http://www.deniscreissels.fr/public/Creissels_Mounole-n.can.val.pat.Basque.pdf

Daniel, Mikhail. 2001. 'Predikativno-argumentnoe jadro [case marking of core arguments]'. In: Kibrik, A.E. (ed.), *Bagvalinkij jazyk* [The Bagvalal language]. Moscow: Nasledie. 369-376.

Dixon, R.M.W. & Alexandra Y. Aikhenvald. 2000. 'Introduction'. In: Dixon, R.M.W. & Alexandra Y. Aikhenvald (eds.), *Changing Valency: Case Studies in Transitivity*. Cambridge: Cambridge University Press. 1-29.

Etxepare, R. 2003. 'Valency and argument structure in the Basque verb'. In J.I. Hualde & J. Ortiz de Urbina (eds.) 2003. 363-426.

Harris, Alice. 1985. *Diachronic Syntax: The Kartvelian Case* (*Syntax and Semantics*, 18). New York: Academic Press.

Haspelmath, Martin. 2010. 'Comparative concepts and descriptive categories in cross-linguistic studies'. *Language* 86(3). 663-687.

Hualde, J.I. & J. Ortiz de Urbina (eds.). 2003. *A Grammar of Basque*. Berlin / New York: Mouton de Gruyter.

König, Christa. 2008. *Case in Africa*. Oxford: Oxford University Press.

Laka, Itziar. 1993. 'Unergatives that assign ergative, Unaccusatives that assign accusative'. In Jonathan Bobaljik and Colin Phillips (eds.) *Papers on case and agreement I*. volume 18. 149-172. MIT Working Papers in Linguistics.

Laka, Itziar. 2000. 'Thetablind case: Burzio's generalisation and its image in the mirror'. In Eric Reuland (ed.) *Arguments and case: Explaining Burzio's generalisation*. 103-129. Amsterdam: John Benjamins.

Lazard, Gilbert. 1994. *L'actance*. Paris: Presses Universitaires de France.

Magomedova, P. 1999. *Čamalinsko-russkij slovar'* [Chamalal-Russian dictionary]. Maxačkala: Dagestanskij Naučnyj Centr Rossiskoj Akademii Nauk.

Magomedova, P. 2003. *Tindinsko-russkij slovar'* [Tindi-Russian dictionary]. Maxačkala: Dagestanskij Naučnyj Centr Rossiskoj Akademii Nauk.

Magomedova, P. 2004. *Bagvalinsko-russkij slovar'* [Bagvalal-Russian dictionary]. Maxačkala: Dagestanskij Naučnyj Centr Rossiskoj Akademii Nauk.

Magomedova, P & I. Abdulaeva. 2007. *Axvaxsko-russkij slovar'* [Akhvakh-Russian dictionary]. Maxačkala: Dagestanskij Naučnyj Centr Rossiskoj Akademii Nauk.

Magomedova, P & Š. Xalidova. 2001. *Karatinsko-russkij slovar'* [Karata-Russian dictionary]. Maxačkala: Dagestanskij Naučnyj Centr Rossiskoj Akademii Nauk.

Mounole, Céline. 2011. *Le verbe basque ancien: étude philologique et diachronique*. PhD thesis. University of Bordeaux & University of the Basque Country.

Oyharçabal, Bernard. 2007. 'Basque light verb constructions'. In Joseba A. Lakarra & José Ignacio Hualde (eds.) *Studies in Basque and Historical Linguistics. In Memory of R. L. Trask*. Bilbao: Diputación Foral de Gipuzkoa - Gipuzkoako Foru Aldundia & Universidad del Pais Vasco - Euskal Herriko Unibertsitatea, Bilbao. 787-806.

Rezac, Milan. 2008a. 'The syntax of eccentric agreement: The person case constraint and Absolutive displacement in Basque'. *Natural Language and Linguistic Theory* 26. 61-106.

Rezac, Milan. 2008b. 'The forms of Dative displacement. From Basauri to Itelmen'. In Xabier Artiagoitia and Joseba A. Lakarra (eds.) *Gramatika jaietan: Patxi Goenagaren omenez*. Universidad del País Vasco / Euskal Herriko Univertsitatea. 709-724.

Saidova, P. 2006. *Godoberinsko-russkij slovar'* [Godoberi-Russian dictionary]. Maxačkala: Dagestanskij Naučnyj Centr Rossiskoj Akademii Nauk.

Samvelian, Pollet. 2012. *Grammaire des prédicats complexes: Les constructions nom-verbe*. Paris: Lavoisier.

Tsunoda, Tasaku. 1985. 'Remarks on transitivity'. *Journal of Linguistics* 21/2. 385-396.