

## **-a and bat Basque articles and recent contact theories**

Julen Manterola

► **To cite this version:**

Julen Manterola. -a and bat Basque articles and recent contact theories. Language contact and morphosyntactic variation and change, Sep 2007, Paris, France. pp.23. artxibo-00350278v2

**HAL Id: artxibo-00350278**

**<https://artxiker.ccsd.cnrs.fr/artxibo-00350278v2>**

Submitted on 18 Oct 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Izenburua - Title**

-A AND BAT BASQUE ARTICLES AND RECENT CONTACT THEORIES

**Egilea – Author**

Julen Manterola Agirre

**Non argitaratzekoa – To be published in**

Language Contact and Morphosyntactic Variation and Change, Vol  
2. Cross-linguistic tendencies in Contact-induced variation and  
change. C. Chamoreau & I. Léglise (eds). CNRS

**Urtea – Year**

2009

**Testuaren izaera - Version**

ZIRRIBORROA – DRAFT

**Sarean non eskuragarri - Available at**

<http://artxiker.ccsd.cnrs.fr/>

# **-A AND BAT BASQUE ARTICLES AND RECENT CONTACT THEORIES**

**Julen Manterola<sup>1</sup>**

**University of the Basque Country**

## **0. Abstract**

The evolution of Basque definite and specially indefinite articles is used to support some recent contact theories. Contact-induced grammaticalization hypothesis is claimed to be a good explanation of Basque indefinite article's diachrony. I will first sketch what we know about Basque articles and how contact is involved in their diachronic evolution; the inaccuracy with which Basque facts have been used in those recent theories will become apparent. First, there are some empirical problems for the aforementioned hypothesis, like the presence of an ancient plural indefinite article *batzu*. Second, some discrepancies arise from the fact that the diachrony of the definite article is neglected. These basic empirical questions will lead to a criticism of some points of contact-induced grammaticalization hypothesis.

## **1. Introduction**

The aim of this article is not to prove whether a certain change has been due to contact or not; rather, I will criticize some claims that, on the shade of some theories of contact, have been made for Basque definite and indefinite articles.

Haase (1992), and following him Heine and Kuteva (2003, 2005, 2006, 2007) have claimed that the indefinite article *bat* 'a' has diachronically developed in Basque induced by such grammatical categories in the romance languages. This issue has not been explicitly targeted by mainstream Basque diachronic linguistics. As for the definite article *-a* 'the', its contact-induced origin is sometimes suggested (Michelena [1978] 1987: 366; Trask 1997: 199), although Haase and works by Heine and Kuteva don't treat it in such terms. Still, as far as I can see, Haase and Heine & Kuteva are the first researchers who taking into consideration current theories on language contact have made an approach to the issue of Basque articles.

The problem with Haase and subsequent works by Heine and Kuteva can be split in two parts: first, when talking about the Basque indefinite article they neglect one important fact, namely the existence of an ancient plural indefinite article *batzu* 'some'; this plural indefinite article is obviously essential for the understanding of the development of the singular indefinite article *bat* 'a'. Second, I identify a problem for Heine and Kuteva's main hypothesis: it would predict the replicated grammatical features to be less developed, but that doesn't apply at all to the case of Basque definite article *-a* 'the', which is much more developed in its grammaticalization path than its counterparts in romance languages.

Thus, the criticism I would like to carry out lies on two basic grounds: it is related to the diachrony of the Basque indefinite and definite articles on the one hand and on the other hand to how Heine and Kuteva fit Basque articles' diachronic development into their theory of contact.

The article is organized as follows: in section 2 I will explain what definite and indefinite articles are in grammaticalization terms. Section 3 will be devoted to the relevant facts we know about Basque articles and their diachrony. Section 4 will focus on how Basque articles have been dealt with in most recent contact theories; finally, in section 5, I will summarize the criticism made in previous sections.

## 2. What we are dealing with. What is an article?

### 2.1 Definite articles

Following Himmelmann (2001), who in his turn relies on Greenberg's seminal work (1978), I will take a fairly strict diachronic view of what a definite article is; it is to that extent that we will be able to compare Basque and Romance definite articles, since one might think they have almost nothing to do with each other (the Basque article on the one side, and Romance articles on the other) from a synchronic point of view. I will illustrate this synchronic difference in behaviour with some examples below (see section 3.1.1).

To put it briefly, a definite article is what we assume to behave as such at a certain point of the diachronic continuum leading distal demonstratives to definite articles first,<sup>2</sup> to specific articles then and finally to noun markers, via grammaticalization, as in the following schema (Himmelmann 2001: 832):

(1) demonstrative → definite article → specific article → noun marker<sup>3</sup>

As a cover term for the grammaticalized elements that derive from a demonstrative, the ones in (1), Himmelmann uses the term D-element. From a strict methodological point of view, then, calling a certain morpheme a definite article is, so to say, a simple convention: we won't find a D-element in one language that behaves exactly the same way as in another.

That is why it is so important to bear in mind a diachronic perspective when analyzing such morphemes, specially when, together with Heine and Kuteva, I will be talking about contact-induced grammaticalization.

There are at least two comments that may be done to this deliberately simple definition:

1. Although not commonly mentioned in the literature, it seems that definite articles may develop from sources other than demonstratives. Frajzyngier (1996) discusses how what he calls "definite markers" have diachronically arisen from items such as *vá* 'hand' in Gidar, and from verbs of saying such as *\*(V)nV* in other Chadic languages; these definite markers seem to display a range of uses analogous to that of "typical", let's say European, definite articles.

Although these cases of grammaticalization provide very interesting data about the diachronic evolution of definiteness marking, I believe they don't affect what I am saying here about the Basque definite article, since we can take for granted without almost any doubt that it belongs to the D-elements continuum schema in (1).

2. The border separating demonstratives from articles and noun markers may sometimes not be clear, although we can use some criteria for this distinction (see Himmelmann 2001). Good examples of how fuzzy can be the borderlines between methodologically established phases of the continuum in (1) are discussions on Chinese and Montagnais, an Algonquian language. They are both said to be "non-article" languages; as for Chinese, works by Huang (1999) and Chen (2004: 1148-1156), among others, suggest that some instances of demonstrative use can be better understood as article-like use; that could perhaps be understood as a hint for an incipient development of a definite article in Chinese. The case of Montagnais, according to Cyr (1993), may result much more extreme, since we may be dealing with a D-element in a very high degree of grammaticalization: in this case, the identity in shape has somehow concealed that D-elements preposed to the noun phrase are in fact articles, in contrast to postposed demonstratives.

These problems do not take place in Basque's grammarian tradition, since Basque's definite article is a well established category. On the contrary, problems may arise from the fact that *-a* 'the' seems to be highly grammaticalized, as we will see in section 3.1 below.

These two possible comments serve to emphasize the specific view by which I approach the diachronic dimension of definite articles. This progressive grammaticalization view is the one taken by Heine and Kuteva (2005) in their contact-induced grammaticalization approach to contact issues. Besides that, it is this very same diachronic point of view the one that allows us to talk about a definite article –it is, crucially, a D-element – in Basque, even though its usage does not resemble too much other western European definite articles'.<sup>4</sup> Usually terminological problems such as calling Basque *-a* "individualizer" instead of definite article are rooted in methodological decisions made when trying to catch the essence of a hypothetic ideal definite article from a strictly synchronic (and most times semantic) point of view.

## 2.2 Indefinite articles

I will also look at indefinite articles in their diachronic dimension, as grammaticalizing items. As already mentioned above, this way of looking at the morphemes is one of the basic points of contact-induced grammaticalization as sketched in Heine and Kuteva's works (see section 4.2.1 below for a basic insight into their hypothesis); I will myself take this view of the issue, since later on I would like to criticize their proposals in these very same terms.

It is well known that cross-linguistically the main source for an indefinite article is the numeral 'one'. As Heine himself points out (1997: 71) "the evolution from lexical to grammatical structure is not discontinuous but proceeds gradually". And looking at the progressive "contextually defined extensions" in the use of the numeral 'one' it is possible to divide these extensions into some descriptively convenient stages. Heine proposes a five-stage model for the diachronic development of the indefinite article, here in (2). I invoke this model mainly because it is the one on which Heine and Kuteva's arguments are based (2006: 104-105), and I feel worth trying to speak in their own terms.

(2) numeral → presentative marker → specific marker → non-specific marker → generalized article

I won't go through the detailed traits of each of these stages; the interested reader is referred to Heine's work (1997: 71-76). I will just recall one of the traits Heine gives for the fifth stage, the generalized article stage: it is in this last stage, he says, where "the use of the article is no longer restricted to singular nouns but is extended to plural and mass nouns, as in the following example from Spanish" (1997: 73). Then he exemplifies how Spanish *uno/una* 'a, one' can be used with plural morphology, *unos/unas* 'some' (the two forms stand for masculine/feminine marking). I understand that he intends this Spanish kind of plural indefinite article to be a characteristic of an ultimate phase on the grammaticalization of an indefinite article. The relevance of this point will become apparent later, as soon as we will see that Basque also has an ancient plural indefinite article.

Besides this diachronic grammaticalization view we take as our basic approach to the issue of indefinite articles, there are some further points worth mentioning about indefinite articles; these points are related to some implicational relationships between definite and indefinite articles. First, it is widely noted in the literature that definite articles develop earlier than indefinite ones, and therefore there are more languages bearing a definite article and not

an indefinite article than vice versa. This is how Heine himself generalizes this observation (1997: 69):

If a language has a grammaticalized indefinite article, it is likely to also have a definite article, while the reverse does not necessarily hold true. Thus, the presence of an indefinite article is likely to be accompanied by that of a definite article, but not vice versa.

Perhaps related to this observation, there is the question of what areal typology could tell us about indefinite articles; I will here further quote Heine (1997: 79):

Thus, one might expect with a certain degree of probability that a given language will have an indefinite article if the neighboring language or language also have one. The older Germanic languages did not have a definite or indefinite article, in much the same way as the ancestor of the modern Romance languages did not. On the other hand, most modern European languages across genetic boundaries have both kinds of article.

Interestingly enough, Basque also has both kinds of articles, although we do not know, at least as long as the indefinite article is concerned, whether it did or not have any prior to contact with indefinite-article bearing languages. I will come back to this point later on next section.

As for the second noteworthy cross-linguistic observation, it seems that languages, rather than extending the former numeral *one* to plural nouns –recall Spanish *unos/unas* ‘some’–, most frequently use alternative strategies to introduce indefinite articles on plural nouns (Heine 1997: 77). This observation regarding the rareness of numeral *one* derived plural articles is also made by Himmelmann (2001: 838), using the same Spanish example. Since languages bearing plural indefinite articles derived from numerals are not cross-linguistically common, it is really meaningful having an area where that kind of items can be found across even not genetically related languages, like Basque and Romance languages. I believe this is an observation to keep in mind.

### 3. Basque facts concerning articles

As a little introduction, let’s sketch how Basque declension works; part of the paradigms here can be checked out in Hualde and Ortiz de Urbina (2003: 173-174). Here I will decline the word *etxe* ‘house’ for its determinerless form, for definite singular/plural and indefinite singular/plural; only four cases will be illustrated here:

Table 1: *Standard inflectional paradigm of etxe, ‘house’*

	- DET	DEF SG	DEF PL	INDF SG	INDF PL
ABS	etxe	etxe-a	etxe-ak	etxe bat	etxe batzu-k
ERG	etxe-k	etxe-a-k	etxe-ek	etxe bat-ek	etxe batzu-ek
DAT	etxe-ri	etxe-a-ri	etxe-ei	etxe bat-i	etxe batzu-ei
GEN	etxe-ren	etxe-a-ren	etxe-en	etxe bat-en	etxe batzu-en

I prefer here to call the nouns in the first column determinerless or bare, rather than indefinite, the term used in the referred grammar above. I keep this term for the nouns bearing an indefinite article in the last two columns. These are maybe just terminological discrepancies; the important point is to know which the morphemes under each label are.

### 3.1 –a: Basque definite article

Basque DOES have a definite article, or at least a D-element as defined in 2.1; thus, from a diachronic point of view the case of Basque’s so called definite article is just one more instance of a grammaticalized distal demonstrative.

Modern Basque’s definite article is –a, a bound morpheme attached to the rightmost element of the whole phrase it modifies; for further information about its behaviour, see Trask (2003: 118-121) and Hualde (2003: 171-177).

#### 3.1.1 Basque definite article’s origin

There are two main forms of the distal demonstrative in modern Basque, depending on the dialect: in western Basque, the distal demonstrative is *a* ‘that’, and in central and eastern Basque is (*h*)*ura* ‘that’. The central and eastern form is said to be a restructured form of a former distal demonstrative, usually reconstructed as *\*(h)a(r)* ‘that’, although no convincing explanation has been given for its exact formation. So, as we can see, western Basque’s distal demonstrative *a* ‘that’ coincides exactly in shape with the definite article –a. Let’s see it in an example from Azkue (1923: 269):

- |   |   |
|---|---|
| (3) a. <i>gizon a</i><br>man that<br>‘that man’ | b. <i>gizon-a</i><br>man-the<br>‘the man’ |
|---|---|

The example of (3b) illustrates the use of the definite article in all varieties of Basque; for surface phonetic variants, see Hualde and Gaminde (1998).

The whole system of demonstratives is reconstructed as having an initial sound, usually an initial aspiration, and that’s why in standard Basque their normative form is *hau* ‘this’, *hori* ‘that’ and *hura* ‘yonder’. This is how it is still uttered in some north-eastern varieties of the language.

In Middle Ages documents there are some instances of the article still bearing the aspiration: *Udalha*, *Adurzaha* (Manterola 2006: 676). These aspiration bearing instances of the D-element are indeed very close to what it has been reconstructed as *\*(h)a(r)*, and they confirm the common opinion about its demonstrative origin (Azkue 1923: 269; Michelena [1971] 1987: 146; Trask 1997: 199).

Summarizing, Basque’s definite article –a perfectly fits the D-element characterization in (1). It is in these terms that we can keep on calling Basque –a a definite article; so has it been called in traditional Basque linguistics. These would be, I believe, the same terms Heine and Kuteva themselves would agree with.

I believe it has been worth clarifying this point about Basque definite article’s origin. It has even been claimed that “one may argue that –a is not really structurally equivalent to definite articles in SAE languages” (Heine and Kuteva 2006: 32). It is true that from a strictly synchronic point of view, Basque’s and neighbouring languages’ definite articles have not too much to do; but when Heine and Kuteva (2003, 2005) talk about model and replica features, they are not thinking, as long as I understand them, in terms of strict synchronic grammatical or structural equivalence, but rather of “grammaticalization path’s equivalences”, so to say. To this extent, inasmuch as it is an instance of the grammaticalization path sketched in (1), Basque definite article is a straightforward parallel to that of romance languages.

The only difference lies on the broader use Basque speakers make of it; as said by Trask “[t]he label ‘definite article’ is misleading, since this article is of much broader use than the English definite article” (Trask 2003: 119). It can be used even in predicates and existential sentences (4), as well as in the citation form of nouns and adjectives. We are so

dealing with a greenbergian stage II towards stage III article (Greenberg 1978: 62-74), towards the rightmost edge of the continuum in (1).

- (4) a. *Ardo-a badago.*                      b. *Ibai irakasle-a da.*  
       wine-the<sup>5</sup> there.is                      Ibai teacher-the is  
       ‘There is wine.’                            ‘Ibai is a teacher.’

We can also find this D-element, *-a*, in adjective predicates:

- (5) a. *Nerea neska jatorr-a da.*  
       Nerea girl nice-the is  
       ‘Nerea is a nice girl.’
- b. *Nerea eta Maider neska jatorr-ak dira.*  
       Nerea and Maider girl nice-the.PL are  
       ‘Nerea and Maider are nice girls.’

Later on it will become apparent why do I give the plural example in (5b), since some interesting contact based reasons have been proposed for these plural predicates (Irigoién 1985: 129). I can’t offer here a complete description of *-a*’s use; in fact, an exhaustive study of its use across dialects and history is still lacking in Basque. For further information about dialectal and historical variation on this D-element’s use, as well as for some possible semantic and functional explanations of its spreading and behaviour, see among others Álvarez (1977), Artiagoitia (1998, 2002) and Etxeberria (2005: 167-250).

### 3.1.2 Basque definite article *-a* and contact

So far, I have made two points about Basque definite article *-a*: first, it has a demonstrative origin, and second, it has gone further in the articles’ grammaticalization path than romance languages. I will now make two further points related to what it has been said in traditional Basque linguistics regarding this *-a* article and contact issues:

1. It has usually been assumed to have arisen due to contact with Late Latin and incipient Romance (Michelena [1978] 1987: 366). In favour of this hypothesis there are at least two noteworthy reasons: languages typologically akin to nowadays Basque (agglutinative, SOV, postpositional) do not usually have a definite article (Himmelmann 1998: 350; Plank and Moravcsik 1996: 205), so contact appears to be an appealing explanation of its existence in Basque. On the other side, Basque seems to have started developing its definite article at the same time its neighbouring languages did (Lapesa 1961; Epstein 1994); that would mean Basque is simply one more example of a widespread Western-European areal phenomenon. The still aspirated instances of the Middle Ages we have seen above point towards this dating of Basque article’s emergence.<sup>6</sup> The issue of the different relative order (Basque *noun + article* versus Romance *article + noun*) has never been addressed.

I believe this contact scenario could be a really suitable one, since the development of definite articles seems to be an areal event of Middle Age’s western Europe (Haspelmath 1998) but we always have to keep clear the borders between our beliefs and our empirically or theoretically based certainties. Until the moment, no extended study has been carried out comparing, inasmuch as data are available, the parallel development of Basque and romance articles.



2. When talking about contact and articles, another issue has deserved some attention in the Basque linguistics tradition: one of the reasons that has been claimed (Irigoien 1985: 129) as responsible for the spreading of the singular and plural definite articles *-a/-ak* is the need of making clear a distinction otherwise non-existent in the language, namely the morphological marking of singular/plural distinction. Since it was in the articles the only place where that distinction was overtly encoded, it seems that, “forgetting” about their definiteness load, both articles (singular *-a* and plural *-ak*) have spread under the model of Romance singular/plural overt morphology (cf. Spanish *cama/camas* ‘bed/beds’).<sup>7</sup> It could then be said that the overt singular-plural distinction in nouns and adjectives has grown at the expenses of the definite singular and plural articles.

The example in (5b) above, as long as we are still waiting for an exhaustive study on this topic, could be an example of this; the Spanish counterparts of the sentences in (5) would be the following:

- (6) a. *Nerea es una chica agradable.*  
 Nerea is a girl nice  
 ‘Nerea is a nice girl.’
- b. *Nerea y Maider son chica-s agradable-s.*  
 Nerea and Maider are girl-PL nice-PL  
 ‘Nerea and Maider are nice girls.’

We can see here that the bare plural predicate *chicas agradables* ‘nice girls’ bear a plural marker, the bold *-s* at the end of both noun and adjectives, a marking modern Basque would have replicated using its phrasal articles. The same facts stand for Trask’s observation that “[...] *ura* may correspond either to ‘water’ or to ‘the water’, and *umeak* may correspond either to ‘children’ or to ‘the children’.” (2003: 121).

Thus, even though the article might have arisen through contact, its spreading path has maybe not strictly followed the typical grammaticalization forces usually assumed for (1). Here it turns out that another contact factor could be related to this spreading, a factor with no direct relationship with the emergence of a definite article, namely the need of an overt morphological distinction of singular and plural already present in the neighbouring languages; this might have played a crucial role on the spreading of the article, in that it is only in articles where singularity and plurality was overtly marked.

### 3.2 *bat*: Basque indefinite article

Basque has an indefinite article, *bat*, which has exactly the same shape as the numeral meaning ‘one’. It is thus common opinion that Basque’s indefinite article has its origin on the numeral. As far as I know, there is no extensive study of its modern use, or how it has evolved last centuries and across different dialects. It is also common opinion that its use is much more restricted than in Romance languages (Trask 2003: 122).

I will offer here a single example, in order to show minimally how *bat* ‘a’ works in contrast to the “definite” *-a* ‘the’. These examples, reminiscent of Givón’s (1981: 36), are both translated with the English indefinite article *a*:

- (7) a. *Azeri bat ikusi dute herrian.*  
 fox a seen have in.town  
 ‘They have seen **a (certain) fox** in town.’

- b. *Azeri-a ikusi dute herrian.*  
fox-the seen have in.town  
‘They have seen **a fox** in town (not a **wolf**).’

We see that *bat* ‘a’ (7a) is used as a specific marker (Heine 1997: 72-73), exactly like in Givón’s street Hebrew’s *-xad*. The noun phrase bearing *-a* in (7b), given the appropriate context, can be interpreted in terms of kind reference; its street Hebrew counterpart would be a bare noun. As far as my language instinct does not betray me, these would be instances of central Basque. There might be (in fact there are) differences across dialects and speakers. However, no in deep study on the different values of *bat* has been carried out up to now.

In French and Spanish both sentences would also be translated with *un*, the indefinite article diachronically issued from the numeral. It is widely noticed in the literature that Basque’s *bat* use is much more restricted than its romance counterparts’. Looking at data like those in (7) one might wonder whether there are other reasons besides the ones proposed by Heine and Kuteva; I will come back to this question later, in section 4.2.3.2.

As for the oldest evidence of the existence of an indefinite article *bat* in Basque, we can only say that it appears in every text from different dialects in the XVI century; that makes us think it is not a recent innovation. Its presence in every dialect could make us think of the age of the ancient common Basque, around the fourth or fifth century; but we have no examples like those we had for the definite article *-(h)a* in the Middle Ages. The kind of corpus we have at our disposal in the Middle Ages – mostly person and place names inserted in Latin or romance texts– does not make probable finding instances of indefinite articles. We simple CANNOT know how old *bat* is in its indefinite article role.

### 3.2.1 *batzu*: the Basque plural indefinite article

Basque crucially has a plural indefinite article *batzuk* ‘some’, morphologically based on the numeral/indefinite article *bat* ‘a, one’. Leaving aside the final *-k*, a newer addition to the older *batzu*, we can dissect it as *bat + zu*. This *bat* is already known for the reader from section 3.2 whereas *-zu* is a collective suffix which is not anymore productive in modern Basque; indeed, its productivity, as long as we can track it back, was already decreasing in the Middle Ages, as shown by Michelena ([1971] 1987:147). This *batzu* is also common to all historical dialects,<sup>8</sup> and it has been present in the records since the very beginning of Basque’s historical period in 1545. This would most probably mean that this *batzu* is at least 1000 years old, or from a time where plurality was marked in ways other than using the articles, as seen in 3.1.2; therefore, one of the possibilities is that it already existed at the time of the ancient basque koiné 1500 years ago (Michelena 1981). Of course, other possibility is that it has later spread from one dialect to other. Michelena himself thinks of it in terms of replication of Spanish *unos/unas*, a hypothesis that could also be possible ([1971] 1987: 148).

Again, we simply don’t know, and maybe CANNOT know.

The important point here for the subsequent discussion is its unquestionable antiquity, as shown by its morphology (the old collective suffix) and especially by its presence in all dialects.

### 3.3 *Basque articles and contact. Summary*

The main guess could be that both articles arose roughly at the same time during the Middle Ages. But there are some caveats to be made on this guess.

As for indefinite articles, the data we have at hand don’t shed any light about its status in the Middle Ages. We simply cannot demonstrate its existence or non-existence prior to Middle Ages contact; it is above all an empirical issue for which no data is available. Put it

otherwise: we have no evidence of a stage of the language where the indefinite article did not exist. And still we have the intriguing ancient plural indefinite article.

The situation for the definite article could seem to be clearer, since even as late as in the XI<sup>th</sup> century there are instances of aspiration bearing articles; that could mean it started grammaticalizing quite “late”. As an alternative hypothesis, however, one could think that it is possible, although maybe not probable, that it started grammaticalizing earlier in the Middle Ages or even before, keeping the aspiration for a longer time. My position in this regard is that, as long as no strong counter evidence appears, the medieval character of the definite article is the least costly hypothesis.

It is worth keeping in mind that these hypotheses have at best the status of most probable guesses; if we are to answer to the question “what was the situation before contact?”, we should make a clear distinction between what we are allowed to consider as empirical evidence, and which are the hypotheses we are assuming.

As a final little note on determinerless nouns, first column in Table 1, what Haase calls transnumerals, it is worth recalling that these forms were much more widely used 500 years ago in every dialect: the most straightforward guess is that the old Basque noun phrase had no overt morphological mark for plurality (besides probably some collective particles) nor for definiteness (Lafon 1954). Eastern dialects, especially Zuberoan and Roncalese, show a rather interesting archaic character in this regard.

## **4. Basque articles and recent literature on contact**

### *4.1 Haase, contact and Basque articles*

It is worth going through Haase’s (1992) analysis on Basque articles, since it is the first study intending to be a general survey on Basque from the point of view of modern contact theories. This task turns out to be unavoidable when we consider that important studies on contact, such as Heine and Kuteva’s, rely almost exclusively on his work as far as Basque is concerned. I will first sum up in subsection 4.1.1 his position regarding the definite article, and then, in subsection 4.1.2., what he says about the indefinite article.

#### *4.1.1 Haase’s Basque definite article*

Haase devotes some 4-5 pages (1992: 53-58) to talk about what I am calling here definite article. He limits himself to sketching a brief description of its use and non use, a description that we can find in Lafitte (1944).

He discusses instances of the definite article in predicative sentences, and following Iturrioz (1985) he states that the Basque definite article *-a* is not really such, but some kind of individualizer. Iturrioz’s analysis of the *-a* morpheme, as far as I can follow his main point, is just a synchronic account of the amazing wide range of uses of the *-a* morpheme.

The inaccuracy of Haase’s description heavily biases Heine & Kuteva’s view on the Basque definite article. I will now list the points I feel lacking in Haase’s analysis:

1. Most importantly, he says nothing about its origin; as I already said in section 3.1.1, its characterization as a D-element is widely accepted in the literature.
2. He doesn’t mention anything about the hypothesis according to which the Basque *-a* morpheme arose through contact. This hypothesis is often mentioned and widely assumed in Basque linguistics (see 3.1.2). He does not mention either anything about the possible role of Romance languages’ singular/plural overt distinction for the spread of the Basque definite

article. That would be another contact issue regarding the Basque definite article Haase does not take into account.

As “minor” points, but essential when one is talking about contact over centuries we can note two further issues:

3. He says nothing about dialect variation in the use of *-a*. It is maybe worth mentioning this, since the immediately contiguous dialect to the one he is analyzing, Zuberoan, shows quite an extensive non-use of the article in contexts where most dialects would use it; most interestingly, the nature of this contiguous dialect’s behaviour has been attributed either to archaism or to French contact (Azkue 1923: 265; Álvarez 1977).

4. He does not deal at all with historical variation. There are many remarks on the progressive extension of *-a* (Lafon 1954; Michelena [1970] 1987: 293, 1978).

So far, these were the facts concerning the Basque definite article Haase doesn’t mention in his 1992 work. We can extend the list pointing at some wrong analysis or methodological gaps when he treats the definite article *-a*:

1. He analyzes the noun phrase of his sentence in (134) (1992: 55), here in (8) as if it was an instance of what he calls transnumeral. I give my own English glosses:

- (8) *Hemen badira jende xahar bat-zu.*  
here there.are people old one-PL  
‘Here there are **some old people.**’

What Haase, following Iturrioz, calls transnumerals would correspond to the determinerless or bare nouns in the first column of Table 1. Since the phrase at stake is *jende xahar batzu* ‘some old people’, and it clearly bears a plural indefinite article *batzu* modifying the noun or adjective phrase, it is wrong labeling it as transnumeral. Furthermore, Basque grammars never mix transnumeral declension with indefinite articles bearing nouns’ declension (see Hualde and Ortiz de Urbina (2003: 118-136) for a recent one).

2. As an additional remark, it’s worth recalling that Haase says nothing about the history and development of definite articles in romance languages, a debate with wide literature; this gap could be, to some extent, understandable, since he is not aware that Basque *-a* is what he would call an *Übersetzungsäquivalent* to romance definite articles.

#### 4.1.2 Haase’s Basque indefinite article

Haase devotes 1-2 pages to the indefinite article *bat* ‘one’ (1992: 59-61, 71). He aims at demonstrating that it arose due to contact but, in my opinion, he states the hypothesis without enough empirical support (Haase 1992: 59):<sup>9</sup>

Der baskischen Transnumeral-Singular-Plural-Opposition steht in den romanischen Kontaktsprachen die Definit-Indefinit-Opposition gegenüber. Hierbei entspricht der indefinite Artikel dem Zahlwort ‘eins’. **Im Sprachkontakt wird das baskische Zahlwort ‘eins’ ebenfalls zum unbestimmten Artikel [...]**

Anders ausgedrückt: *bat* und frz./gask. *un* sind im Bereich der Zahlwörter *Übersetzungsäquivalente*. Wie in anderen Fällen [...] kann sich nun der Funktionsbereich von *bat* auf alle die Fälle ausbreiten, in denen in den Modellsprachen *un* gebraucht wird, also auch auf die Signalisierung von Indefinitheit.

[The Basque transnumeral-singular-plural opposition stands against the definite-indefinite opposition of Romance contact languages. In that sense the indefinite article corresponds to the

numeral ‘one’. **In a language contact situation the Basque numeral ‘one’ will likewise become an indefinite article [...]**

In other words: *bat* and French/Gascon *un* are in the scope of the numerals *translation equivalents*. As in other cases [...] the range of functions of *bat* can be extended to all the cases, in which *un* would be used in the model languages, therefore also to mark indefiniteness. (Emphasis mine JM)]

After having stated this, he proceeds with two more examples through which he shows that *bat*, the numeral ‘one’ moving towards an indefinite article, has been extending its semantic meaning. He first gives a sentence from a 1782 work, and then another one from the first printed book in Basque (1545), here in (9) (his (161-162), (1992: 60)).

- (9) *balia dikezit        senhar        gaixto bat*  
can be.for you    husband    bad    a  
‘I can be as good as a bad husband for you’

He adds a comment regarding the use of the indefinite article *bat* in this sentence:

Der Gebrauch von *bat* könnte durch das Verb *baliatu* ausgelöst worden sein. Im Keim zeigt sich aber schon die **im Sprachkontakt katalysierte Entwicklung**.

[The use of *bat* could have been triggered by the verb *baliatu*. However, at source it appears to be an evolution catalyzed by language contact.]

I don’t feel especially convinced, with a single example from a single language, about the statement claiming that this *bat* use was triggered by language contact. I don’t mean language contact plays no role (in fact, I believe it may have played a determinant role), but I would expect, in order to support this claim, a much deeper analysis, with examples of as many old texts as possible, comparing them to data from other dialects, periods and model languages. And even then, after having “squeezed” our data as much as they allow us to, we sometimes have to admit we can not go any further. Anyway it might be, a deep knowledge of old texts and dialects comes always first, as I have just suggested. These are our tools, we cannot skip over them.

Summarizing, I would say that two examples, dating from 1545 and 1782, aren’t enough support for the claim that Basque *bat* and French and Gascon *un* are translation equivalents nor for directly inferring the direction of an alleged contact-induced change.

Besides these poorly exemplified statements, there is a crucial silence on other aspect of the indefinite article: its plural *batzu* ‘some’. Since he just omits this fact, we cannot know whether he would also attribute its existence to contact. It is worth recalling that this plural indefinite article *batzu* has to be quite ancient (see section 3.2.1), or at least older than Haase’s description together with Heine’s grammaticalization scale of indefinite articles in (2) would let us think.

There is one further statement by Haase that deserves some comment (Haase 1992: 61); this is his last statement before he proceeds to deal with case and postposition systems:

Das baskische Determinationssystem, das auf der Opposition zwischen Transnumeral, Singular und Plural beruht, ist – wie wir gesehen haben – destabilisiert worden. Zum einen wird die Unterscheidung von nicht-individualisiertem und individualisiertem Prädikatsnomen aufgegeben, zum anderen wird das Zahlwort für ‘eins’ nach romanischen Vorbild zum indefiniten Artikel, der anstelle des Individualisierers eintreten kann.

[The Basque system of determination, that stands on the opposition between transnumeral, singular and plural, has been – as we have seen it – destabilized. On the one hand, the difference between non-individualized and individualized noun-predicates is abandoned, on the other hand

the numeral for ‘one’ becomes under romance model an indefinite article, which can take the place of the individualizer.]

It is difficult for me to understand which span of the time axis is he taking as a basis each time: sometimes it seems he is talking about very recent changes in Basque. An example of this could be his sentence in (166), from his own fieldwork, given just before this last quote; I repeat it in (10) with my own glosses:

- (10) *tokero bat zen*  
driver a was  
‘(S)he was a cattle-driver’

This, indeed, could be a very nice example of how *bat*, on the model of Romance languages, since in the dialect he takes examples from this profession noun predicate would usually bear no determiner, and in western dialects it would take *-a*.

But based on Haase’s data we do not know how ancient this kind of *bat* instances are; nor can we know it based on the sole example Haase provides in his analysis. We don’t know either which is exactly the romance model he has in mind. It doesn’t seem, in his favour, that Haase intends this sentence to be more than an example of current contact induced use of *bat*. As long as he would take this kind of examples as evidence for a recent destabilization of what he calls Basque transnumeral-singular-plural opposition, I could, perhaps, agree with him to some extent.

Some other times, and always sticking to the same last quote, one feels like Haase is talking about quite old changes: when he says that Basque’s *bat* became an indefinite article under the model of romance languages, one can assume that Haase is aware of the relative antiquity of *bat* in indefinite article uses. If this is so, then there is a problem when he treats all contact induced changes and analysis on language systems/oppositions of different chronology together. I will try to explain this briefly.

The Transnumeral-Singular-Plural opposition he takes for granted as the ancient and original of Basque’s determination system, the one represented by the three leftmost columns in Table 1, is not such.

As discussed in 3.1.2 this nowadays Basque’s opposition can not really be the “original” Basque one:<sup>10</sup> what Haase calls singular and plural bear in fact a demonstrative based definite article. We have also seen that they arose at the same time in the Middle Ages as in some other western European languages, most probably in an areal configuration. The problem is, again, that Haase does not take his “individualizer” as an instance of the D-elements grammaticalization path (section 3.1.1).

Let’s assume now that roughly at the same time, in the Middle Ages too, an indefinite article arises (recall our ignorance about its exact emerging date, section 3.3). So what we in fact have is that the alleged entering of *bat* in Basque’s system could well go together with the arising of the singular and plural articles and the configuration of modern Basque’s bare-noun vs. definite articulated-nouns, (Haase’s transnumeral-singular-plural opposition). We thus see that the indefinite article *bat* has not been destabilizing any former transnumeral-singular-plural opposition; this latter opposition has most surely been growing together with the further grammaticalization of the indefinite article *bat*.

All my comments here, I believe, make much more difficult to understand what Haase means when he treats the Basque transnumeral-singular-plural system as opposed to Romance definite-indefinite. Besides that, recall again that the singular-plural opposition might have grown “parasitizing” the definite-indefinite one, as suggested in section 3.1.2.

As a last comment on the quote above regarding the transnumeral-singular-plural opposition Haase takes as originally Basque, we can recall what I said in section 3.3. Most

surely, the “original” Basque, let’s say 1600 years ago, was a language with no Romance fashioned overt morphological marking singular vs. plural, nor any overt morphological definiteness marking.

Whatever we might think about how Haase deals with contact and the indefinite article *bat*, I believe some of the spreading processes of the indefinite article could be accounted for in terms of contact; we should, nevertheless, start placing things in their correct relative chronology, paralleling them with romance articles’ development. This is actually a basic task that remains undone: as a prospective work, Basque linguistics should someday go through the description of the indefinite article’s uses in historical data from 1545 across different dialects. Change has to be asserted supported by as many examples as possible, coming from different dialects and historical periods.

Unfortunately, Haase’s contribution, with a discussion on four or five examples of *bat*, is not helpful for the accomplishment of this task. As I have tried to show here, there are some points in his reasoning that should be taken cautiously and others that could be better understood with a wider knowledge of Basque’s diachrony.

#### *4.2 Heine and Kuteva’s model of contact and Basque articles*

In this section I will first offer a brief sketch of some generalizations Heine and Kuteva make for language contact situations (section 4.2.1.); then I will focus on how they have treated Basque indefinite and definite articles (4.2.2).

##### *4.2.1 Generalizations on contact-induced grammaticalization*

One of the basic ideas of contact-induced grammaticalization as explained by Heine and Kuteva is that change is gradual rather than abrupt. Speakers of the replica language activate a pattern of their own language, the one corresponding most closely to the model; this way they develop a structure that is equivalent to the one in the model language. This pattern develops, eventually grammaticalizing into a new full-fledged grammatical category, similar to that of the model language (Heine and Kuteva 2005: 121). Thus, although initially lacking a category structurally equivalent to that of the model language, a certain pattern of the replica language may undertake a grammaticalization path analogous to the one the model language may have undertaken previously. To this extent, it is legitimate thinking that the similarity between replica and model language lies especially on the fact that they share the same grammaticalization path for their parallel structure.

Other of the main points Heine and Kuteva make about contact-induced grammaticalization is well summarized in next quote (2005: 101):

[W]herever there is sufficient evidence, it turns out that the replica construction is less grammaticalized than the corresponding model construction [...] in the initial stage of grammaticalization, the new category tends to be ambiguous between its literal and its grammaticalized meaning, it tends to be confined to few contexts, and its use is optional. [...] Such properties are commonly encountered in replicated categories.

Thus, according to them, when a language “copies” a certain feature or category, it does it in this progressive way, somehow starting a grammaticalization process of its own, paralleling the one of the model language.

This hypothesis is a very appealing one, since it allows us to predict “backwards”, on the basis of the grammaticalization degree of two features, which language has been the model and which one the replica (Heine and Kuteva 2005: 120):

[I]t seems possible to determine in a situation **where no diachronic information is available** which is the model and which is the replica category. (Emphasis mine JM)

Of course, the authors themselves are aware of the limits of this so to say predictive power and admit that if the contact situation lasts long enough, both categories, model and replica, may eventually “become structurally indistinguishable” (2005: 120).

The risk, as I see it, is falling in a circular argument and determining the contact relationship of two languages on the basis of this hypothesis; an in deep knowledge of the diachrony of the languages at stake should always come first.

These few hints on Heine and Kuteva’s approach to contact-induced grammaticalization will suffice to understand the discussion in next sections. Next section will be useful as a reminder of how cautious we have to be when we supposedly lack diachronic information. Basque data, which at first sight seemed to fit this hypothesis, turn out to be even contrary to it if we look at them with no bias of any kind of theoretical expectations.

#### 4.2.2 Basque definite article in Heine and Kuteva’s works

As far as I have been able to track their works, Heine and Kuteva discuss explicitly the definite article only in one book (2006: 32). Let’s quote them:

[T]he primary function of the ‘definite article’ *-a* is to individualize referents, and these referents can be, and not uncommonly are, indefinite or even non-specific [...] that Basque has a definite article can be justified on the grounds that *-a* is more likely to mark definite than indefinite reference. However, **one may argue that *-a* is not really structurally equivalent to definite articles in SAE languages**. In this case, a taxonomic conclusion that one could draw from the observations made is that, rather than having a definite but no indefinite article, Basque has an indefinite but no definite article –hence, quite the opposite of what a discrete-categorization approach of the kind employed by the typologists cited suggests. (Emphasis mine JM)

Although they don’t make any strong statements about the definite article in Basque, it is clear that relying exclusively on Haase’s work has heavily biased their view on what actually has been going on last 1000 years.

It is true, as they point out, that *-a* can be used for non-specific reference (see (6b)), and to that extent one could argue it is something else than a definite article; again, this only depends on what we understand by definite article.

But I don’t believe that being or not structurally equivalent to other languages is relevant at all for the point Heine and Kuteva want to make. Indeed, this is even contrary to the terms they themselves are proposing for their own approach. For what has to be equivalent between the languages we compare is the grammaticalization path of the relevant feature in each language, as suggested in section 4.2.1. Moreover, at least in what definite articles are concerned we would barely be able to find such a structural equivalence across languages, depending of course on how we define being structurally equivalent.

As already seen in previous sections (3.1.1), Basque’s so called definite article fits perfectly the grammaticalization path leading from demonstratives to articles; and Romance languages’ articles also do, inasmuch as many of them have a Latin *ille* origin. To this extent I guess Heine and Kuteva would have to admit that we CAN compare Basque’s and Romance languages’ definite articles. Moreover, they have already been compared before in the literature, and Basque definite article has been claimed to have arisen in a contact situation (section 3.1.2).

Yet, a problem would immediately arise for Heine and Kuteva’s reasoning: Basque *-a* is used in a more extended way than its romance counterparts. As long as Heine and Kuteva would have cared about it, this relative relationship between the grammaticalization degree of Basque and romance articles would not fit into their expectations: their working hypothesis is



that Basque, as the replica language, should have its replica feature much less grammaticalized than the model languages.

As far as I can see, there would be two logically possible solutions to this situation if we still want to stick to Heine and Kuteva's hypothesis: either (1) Basque is the model language whereas Romance languages are the replica languages or (2) it might be the case that the contact situation has been so long that former relative grammaticalization degrees between replica and model features have been blurred by time. Let's now try to develop what the problems are for each of these two possibilities:

1. Should we say that Basque is the model language when dealing with the definite article, but the replica language when dealing with the indefinite one? As we will see below, and as already suggested before, according to Heine and Kuteva Basque indefinite article's development fits perfectly with the characterization of it as a replica feature. Is this a problem if we want to say that the Basque definite article is in its turn a model feature for Romance languages? I guess the most straightforward answer to this is "yes, it is".

If we would be taking this explanation path, we would be deciding which one is the replica and which the model language not on the basis of actual sociolinguistic data (which we lack), but on the basis of what is more convenient to our hypothesis. In order to avoid this and further methodological problems I think we should disregard this first logical possibility.

What the discussion carried out for this first logically possible solution indirectly suggests us is that we are maybe not allowed to decide, relying on the relative grammaticalization degree of the relevant features, which one is the replica and which one the model feature. This should make us reconsider the position Heine and Kuteva take regarding the Basque indefinite article *bat*; we will come back to this issue in next section (4.2.3.1).

2. The second possibility is indeed a solution Heine and Kuteva themselves take into consideration, although not specifically for the Basque case. Look at this quote (Heine and Kuteva 2005: 265)

[O]ne caveat with regard to this generalization: given enough time, replica categories can develop in the same way as their models [...]

Again, if we are to follow this possible solution, a problem mirroring the one we sketched for the first solution arises: if the contact situation has been so long that the replica category (definite article *-a* in our case) has developed to the same degree (and beyond, in this case) as its models, what should we say about the indefinite article *bat*? Is not the contact period equal for both definite and indefinite articles? As I suggested above (section 3.2), Basque's indefinite article could be at least as old as the definite one. Another possibility, of course, is that in our alleged model language(s) definite and indefinite articles' grammaticalization happened at different paces or times. There is maybe a deeper question floating over these considerations, and it is the one regarding the relative speed on which language change happens; I won't enter this discussion now.

As briefly sketched here, some problems arise when we also take the definite article's development as something that should be explained in the same principled way Heine and Kuteva propose for other features. Furthermore, when we analyze it together with the indefinite article *bat*, the only feature of both definite and indefinite articles they try to account for, new problems surface for their hypothesis. In next section I will discuss in more detail the indefinite article.

#### 4.2.3 Basque indefinite article in Heine and Kuteva's works

Basque indefinite article *bat* is discussed or mentioned in four works by Heine and Kuteva (2003: 556-557, 2005: 101, 247, 2006: 30, 132, 246, 2007: 327). Here I will only offer some relevant quotes, since the ideas presented in each of these works do not differ too much from each other.

#### 4.2.3.1 *bat* indefinite article as a replica feature

Heine and Kuteva take for granted the acquisition of the indefinite article via contact, although we have no evidence of a time where Basque lacked such a category. See this quote (Heine and Kuteva 2003: 556):

As a result of this contact, Basque speakers introduced a **category which they did not have previously**, namely an indefinite article. (Emphasis mine JM)

We can find other claims that go in the same vein throughout their different works (Heine and Kuteva 2005: 247). Here I offer another quote from their last work (Heine and Kuteva 2007: 327):

[I]n the earlier history of the Basque language there was no indefinite article, while the surrounding Romance languages Spanish, French, and Gascon had indefinite articles. As a result of centuries of close contact with these Romance languages, speakers of Basque grammaticalized their numeral for 'one', *bat*, to an indefinite article. [...] as Haase (1992) demonstrates, it was only one out of a large number of instances of grammatical replication that Basque speakers introduced on the model of their dominant Romance neighbor languages [...]

These words are very explicit about the non-existence of an indefinite article in the earlier history of Basque. Of course it depends on how we understand the term "history", but as I have shown in sections 3.2 and 3.3, they are strictly speaking not true: *bat* as an indefinite article appears in all Basque historical records. There are not extensive records of Basque for the time it allegedly would have lacked an indefinite article; again, we simply don't know when it emerged in Basque. As a possibility, as plausible as any other, we should consider that Basque could have had an indefinite article prior to contact with Romance languages.

Of course, we know (see section 2.2) that languages tend not to have only an indefinite article, so one could think, as quite a probable hypothesis based on cross-linguistic tendencies, that Basque did not have an indefinite article before the definite one emerged, allegedly in the Middle Ages. But we must make a distinction between what we decide on the basis of our theoretical assumptions and what we know for sure on the basis of actual data. And we also should keep in mind that languages like Turkish are an exception to that tendency, especially when this language seems to be close to Basque in typological terms (Comrie 2008).

Summarizing: we can't make a decision on this issue, whether Basque had or not an indefinite article before the Middle Ages, depending on which possibility fits better into a given hypothesis: it is first of all an empirical problem, the one of the lack of relevant data, or rather the one of the good use we should make of the data at our disposal.

#### 4.2.3.2 *Gradual (and "delayed") grammaticalization process in replica language*

Heine and Kuteva also focus on the lower degree of grammaticalization Basque *bat* shows as compared to Romance *un* 'a'. This situation would fit their expectations about model and replica features' relative chronology: Basque *bat* is less grammaticalized. See next quote (2003: 556-557):

The grammaticalization of indefinite articles normally proceeds along the following main stages [...] While the French indefinite article has gone essentially through all these stages, the Basque indefinite article has not. [...] While there are incipient uses as a non-specific marker as early as 1545, the grammaticalization as a non-specific article is clearly a recent innovation of Basque.

[...] it has not reached the same degree of grammaticalization as e.g. the corresponding French article [...]

The only relevant quote from their 2005 work shares essentially the same idea (Heine and Kuteva 2005: 101):

[...] and the indefinite article of Basque, replicated on the model of Romance languages, exhibit properties of categories in the early stages of grammaticalization [...] They thus differ from the corresponding categories in the model languages, which both are fully grammaticalized articles.

I don't believe that Basque's indefinite article is less developed than in the alleged model languages exclusively because it is a replica category. Look at the following example of nowadays central Basque:

- (11) a. *Eneko gizon on-a da.*                      b. ??*Eneko gizon on bat da.*  
Eneko man good-the is                      Eneko man good one is  
'Eneko is a good man.'

In its romance counterparts we have most typically *Eneko est un bon homme* or *Eneko es un buen hombre*, with the *un* indefinite article. It becomes clear that there may be other issues at stake: how did the spreading of *-a* affect to the use and further grammaticalization of *bat*? The role the amazing spread of the definite article may have played preventing indefinite *bat*'s use should also be considered.

At the same time, there is a mirroring phenomenon worth bearing in mind: *bat* can be used in such constructions in some texts and varieties, mostly in, this is my impression, eastern dialects. Examples like *Fantasma bat da* 'It's a phantom' by Leizarraga (1571: 334, Mat. XIV, 26) are a good example of it; however, it is also possible to find examples like *Lecu on bat da Escocia* 'Scotland is a good place' in Pérez de Lazarraga (c. 1564: 1204r), a representative of western basque.<sup>11</sup>

The mirroring possible question would then be whether this broader use of *bat* could be related or not to the lower use of *-a* in such dialects. This is a further point lacking detailed study and the guesses I come here with are all waiting for empirical test.

These are questions raised by Basque internal facts of definiteness marking and behaviour of determiners; I think their general theory should also offer solutions to this kind of language specific problems. In their favour, I can here recall Heine and Kuteva's own idea: a long contact period may have blurred the relative grammaticalization degree of model and replica languages; but to keep on playing fair, we should then come back to the problems I sketched in 4.2.2. No clear solution can be found sticking to their terms, at least for Basque facts.

On the other hand, one could think what does mean not being strongly grammaticalized (in Heine's 1997 terms), since Basque has a plural indefinite article *batzu*. Next section will be devoted to this topic.

#### 4.2.3.3 Plural indefinite article *batzu* and their theory

In their 2006 work they mention the indefinite article *bat* three times (2006: 29, 132, 246); they discuss it in the same terms as in their previous works, always relying on Haase's (1992) work. The main bulk of the criticism made for Haase in section 4.1 is valid here, so I won't go through all the detail again. There is, still, a couple of further comments I would like to make on next quote (2006: 246):

[B]ut the grammaticalization as a non-specific article is clearly a recent innovation of Basque. While the Basque article exhibits a high degree of grammaticalization, it is still less grammaticalized than its equivalents in the Romance model languages.

First, the data they offer, those showing that the grammaticalization of a non-specific article is a recent innovation, are not conclusive. I don't mean it is not the case, but Haase's few examples are not enough, in my opinion. This is a further topic for which a detailed analysis, based on as many texts as we may have at our disposal, is still lacking.

Second: what would in Heine's (1997) terms mean that the Basque article is still less grammaticalized? I have recalled many times the existence of an ancient plural indefinite article *batzu* (section 3.2.1). I see two logical possibilities if we want to keep Heine and Kuteva's (2003, 2005, 2006, 2007) and Heine's (1997) view on the issue: if we follow Heine's indefinite article's grammaticalization scale in (2), especially what he says for the last stage, we would have to admit that Basque indefinite article *bat* was very grammaticalized at an ancient time. If this would be the case, Heine and Kuteva's works' basis will be weakened; it would seem that the usual grammaticalization path does not apply to this specific case. As a second logical possibility, one could maybe think that Heine's scale is not correct, since Basque indefinite article *bat* is not so advanced in its grammaticalization path, yet it has an ancient plural *batzu*.

The striking point is, of course, that *bat* as an indefinite article seems to be grammaticalized and not grammaticalized both at the same time. If we feel free not to adhere to these authors' theoretical proposals, we might think that the grammaticalization of the indefinite article, triggered or not by a contact situation, has not followed the typical path proposed by Heine.

Actually, as we have seen before (section 3.2.1), plural *batzu* could indeed have been shaped on the model of romance languages (at least Spanish); the fact that plural indefinite articles based on the numeral are rare (section 2.2), can also make us think of an areal feature.

Many questions come to my mind now, which will inevitably remain open for discussion: what do really strong contact situations mean for grammaticalization scales? We could perhaps answer that the model indefinite plural *unos* 'some' was so strong that made Basque *bat* 'a' skip over some stages in its grammaticalization path. Thus we would not need to reject Heine's grammaticalization path in (2); and we would open a new research line for the interaction between "natural" grammaticalization paths and contact affected ones. This option could perfectly be complementary to Heine and Kuteva's proposals for grammaticalization and contact issues.

From a more local perspective, we could also wonder about the relationship between Spanish language and nowadays French Basque dialects, or what the internal relationship between western and eastern dialects has been or, again, we could also think about the complementary relationship between definite *-a* and indefinite *bat*. General theories should also have a word for these seemingly not so important issues.

#### 4.2.4 Concluding remarks on Heine and Kuteva's hypothesis for Basque

I have tried to show that Basque data as used by Heine and Kuteva don't present a good support for any of their contact hypotheses. Nevertheless, I would like to make clear that contact has most surely played a determinant role configuring Basque articles' character and range of uses. On the one side, although I have not focused on this issue, the amazing extension of the definite article *-a* could be explained by contact: not simply by means of "typical" contact-induced grammaticalization of demonstratives, but rather by another contact effect, the spreading of the overt marking of singular / plural morphology.

On the other side, as for the indefinite article *bat*, it is true that its use has been extended on the model of romance languages, as Trask (2003: 122) notes:

The quantifier *batzuk* 'some, several' [...] is formally a plural of this *bat*. Among some younger speakers, there is a tendency to extend the use of *bat* to calque the much broader use of the Spanish article *un(a)*.

This extension can in fact be a measurable contact induced change, but recent works by Heine and Kuteva (with Haase on their basis) don't address this phenomenon. This fault is not entirely their own, Basque linguistics in general still lacks an in deep study on the topic.

Many interesting questions arise from this probable contact induced extension; I will sketch here a couple of them, as possible guesses:

1. As mentioned before, the indefinite article's extension could have been affected by the widespread use of the definite article; this definite article is the one which most evidently has been spreading its use last four-five centuries when Basque as a whole is considered
2. On the other hand, it might seem that eastern varieties spread much more the indefinite article on the model of romance languages than central and western varieties. One could think that this can be due to the lower use eastern varieties make of the definite article *-a*.

These shortly mentioned possibilities, as I said, are simple guesses; further study is needed in order to corroborate or dismiss them. They are intended to be a simple sample of what kind of research lines we may follow, since they are not addressed by Haase.

## 5. Summary and last remarks

So far, it has become clear that the reality of linguistic facts offers us a much more complex, multi-coloured situation than what Haase's and Heine and Kuteva's proposals would let us imagine.

I have shown that Heine and Kuteva's hypotheses cannot hold for both Basque articles, definite and indefinite, together (section 4.2.2). But the basic problem comes from the fact that as far as Basque is concerned they rely exclusively on Haase's work. There are many other sources with which Haase's seminal work on contact should be checked, in order to get a reliable idea of diachronic, dialectal and contact issues' analysis on Basque.

Dialectal and historical data can not be neglected; proposals coming from theoretical insights should help understanding the history and dialectal variation and distribution of language specific features. But in their approach many issues are not addressed and remain unexplained:

1. It doesn't solve specific problems already identified in Basque diachronic linguistics, like the early existence of the indefinite plural *batzu* (section 3.2.1).
2. It doesn't say anything about contact issues present in the literature on Basque, like the contact induced emergence of the definite article *-a*.
3. There is no mention of the amazing spreading of the definite article *-a*, which seemingly could have been due to contact.
4. Interesting questions, like the relationship between definite and indefinite articles in a contact situation are not targeted.

This whole new view on Basque issues regarding articles forces us to come back to their theoretical claims (section 4.2.1), insofar as Basque facts were supposed to support them:

1. The relative grammaticalization degree of parallel categories is not to be taken as the first approach to contact issues between two languages. When we want to establish their diachronic relationship other questions have to be solved first.
2. When diachronic information is not available we should behave in much a more cautious way; the example from Basque has shown that when we indeed did have better and more

information than the authors (considering both articles together), their predictions were simply not correct.

3. Thus, their caveat in the quote I brought here in section 4.2.2 points towards an interesting direction for further studies, focusing on the time dimension of these developments.

4. New questions arise for their hypotheses, as how different contact effects can conceal each other's typical diachronic paths (see section 3.1.2, about the role of singular/plural marking overt distinction). Looking at how really strong contact affects to the replica language should also be a concern of the theory (see sections 3.2.1 and 4.2.3.3 about the possible contact origin of *batzu*).

All these thoughts, rather than completely ruling their hypotheses out, could maybe help to improve them and to update them to what we empirically know about facts in Basque. They also should be a call for cautiousness, as long as we are dealing with a language with no decisive data in some aspects, and a reminder of how important it is to acquire a good knowledge of the history of the languages involved.

At this point it is worth recalling some aspects of what Thomason (2007) proposes as basic steps that must be taken before a claim of contact-induced change can be considered to be firmly established. As directly tied to the point I was making about the importance of a good knowledge of the history of the languages involved, she gives her fourth and fifth steps to be taken (Thomason 2007: 11-12): on the one hand, we need to prove that the proposed interference features did not exist in the receiving or replica language before it came into contact with the source language; this can be done either by inspecting the documents showing earlier stages of the language, or by examining related languages which can give us hints about how the ancient mother language looked like. And the fifth step, of course, would be proving that the relevant transferred features were present in the source or model language by the time it came into contact with the receiving language. As far as I can judge, none of these basic steps has been accomplished by Haase for Basque definite and indefinite articles, nor have they been checked by Heine and Kuteva; ancient texts are not well analyzed, and variation between dialects (replacing related languages in the case of isolate Basque) has not been studied.

These methodological gaps cause their analyses on Basque contact issues not to be precise. Furthermore, one would expect their hypotheses to shed some light on some language specific problems, so theory and empirical facts would strengthen each other. This has not been the case; Basque has been shown not to be a good starting point for their hypotheses in what articles are concerned. However it might be, Heine and Kuteva's hypotheses have shown that they are testable against new data, and therefore their theory may turn out to be strengthened as these data will, little by little, feed it.

## 6. References

- Álvarez, J.L. 1977. L'emploi de l'indéfini en souletin. *Fontes Linguae Vasconum* 25: 29-54.
- Artiagoitia, X. 1998. Determinatzaile sintagmaren hipotesia euskal gramatikan. *Uztaro* 27: 33-61. [Available at <http://www.uztaro.com/>]
- Artiagoitia, X. 2002. The functional structure of the basque noun phrase. In *Erramu Boneta: Festschrift for Rudolf P. G. de Rijk* [ASJ Uren Gehigarriak XLIV], X. Artiagoitia, P. Goenaga and J. A. Lakarra (eds), 73-90. Bilbo: EHU.
- Azkue, R.M. 1905-1906. *Diccionario vasco-español-francés*. Bilbao: La Gran Enciclopedia Vasca. [Reprint in 1969]
- Azkue, R.M. 1923. *Morfología Vasca*. Bilbao: La Gran Enciclopedia Vasca. [Reprint in 1969] [Available at <http://www.euskaltzaindia.com/euskera>]
- Companý, C. 1991. La extensión del artículo en el español medieval. *Romance Philology* XLIV (4): 402-424.

- Comrie, B. 2008. Basque, Romance, and areal typology: What do we learn from the World Atlas of Language Structures? In *Lenguas en diálogo. El iberorromance y su diversidad lingüística y literaria. Ensayos en homenaje a Georg Bossong*, H.-J. Döhla, R. Montero and F. Báez de Aguilar (eds), 55-63. Madrid: Iberoamericana.
- Cyr, D. 1993. Cross-linguistic quantification: definite articles vs demonstratives. *Language Sciences* 15 (3): 195-229.
- Chen, P. 2004. Identifiability and definiteness in Chinese. *Linguistics* 42 (6): 1129-1184.
- Egiategi, J. 1785. *Lehen liburua edo filosofo huskaldunaren ekheia* [Euskararen Lekukoak 6]. Bilbao: Euskaltzaindia. [Reprint in 1983]
- Epstein, R. 1993. The later stages in the development of the definite article: evidence from french. In *Historical linguistics* [Current Issues in Linguistic Theory 124], H. Andersen (ed), 159-175. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Epstein, R. 1994. The development of the definite article in French. In *Perspectives on grammaticalization* [Current Issues in Linguistic Theory 109], W. Pagliuca (ed), 63-80. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Epstein, R. 1995. L'article défini en ancien français: l'expression de la subjectivité. *Langue Française* 107: 58-71.
- Etzeberria, U. 2005. Quantification and domain restriction in Basque. PhD Dissertation, University of the Basque Country. [Available at [www.ehu.es/hitt/etzeberria.htm](http://www.ehu.es/hitt/etzeberria.htm)]
- Frajzyngier, Z. 1996. On sources of demonstratives and anaphors. In *Studies in anaphora* [Typological Studies in Language 33], B. Fox (ed), 167-203. Amsterdam/Philadelphia: John Benjamins.
- Givón, T. 1981. On the development of the numeral "one" as an indefinite marker. *Folia Linguistica Historica* II (1): 35-53.
- Greenberg, J.H. 1978. How does a language acquire gender markers? In *Universals of Human Languages*, J. H. Greenberg, J. Ferguson and E. Moravcsik (eds), 47-82. Stanford: Stanford University Press.
- Haase, M. 1992. *Sprachkontakt und Sprachwandel im Baskenland: Die Einflüsse des Gaskognischen und Französischen auf das Baskische*. Hamburg: Helmut Buske Verlag.
- Haspelmath, M. 1998. How young is standard average European? *Language Sciences* 20 (3): 271-287.
- Heine, B. 1997. *Cognitive foundations of grammar*. New York, Oxford: Oxford University Press.
- Heine, B. and Kuteva, T. 2003. On contact-induced grammaticalization. *Studies in Language* 27 (3): 529-572.
- Heine, B. and Kuteva, T. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Heine, B. and Kuteva, T. 2006. *The changing languages of Europe*. Oxford: Oxford University Press.
- Heine, B. and Kuteva, T. 2007. *The genesis of grammar. A reconstruction* [Studies in the evolution of language 9]. Oxford: Oxford University Press.
- Himmelmann, N.P. 1998. Regularity in irregularity: Article use in adpositional phrases. *Linguistic Typology* 2: 315-353.
- Himmelmann, N.P. 2001. Articles. In *Language Typology and Language Universals: An International Handbook* [Handbooks of Linguistics and Communication Science 20.1], M. Haspelmath, E. König, W. Oesterreicher and W. Raible (eds), 831-841. Berlin-New York: Walter de Gruyter.
- Hualde, J.I. 2003. Case and number inflection of noun phrases. In *A grammar of Basque*, J. I. Hualde and J. Ortiz de Urbina (eds), 171-186. Berlin, New York: Mouton de Gruyter.
- Hualde, J.I. and Gaminde, I. 1998. Vowel interaction in Basque: a nearly exhaustive catalogue. *Studies in the Linguistic Sciences* 28 (1): 41-77. [Available at [https://netfiles.uiuc.edu/jihualde/www/V\\_sequences.pdf](https://netfiles.uiuc.edu/jihualde/www/V_sequences.pdf)]
- Hualde, J.I. and Ortiz de Urbina, J. (eds). 2003. *A grammar of Basque*. Berlin, New York: Mouton de Gruyter.
- Huang, S. 1999. The emergence of a grammatical category definite article in spoken Chinese. *Journal of Pragmatics* 31: 77-94.
- Iglesias, H. 2007. Observations concernant les récentes critiques et omissions de Joseba Lakarra à propos des recherches d'Héctor Iglesias sur la problématique "basco-ibérique" suivies d'une hypothèse inédite concernant l'inscription de Liria. [Available at <http://artxiker.ccsd.cnrs.fr/>]
- Irigoyen, A. 1985. Euskarazko izen sintagma mugatzailerik gabekoez. *Euskera* 30: 129-139. [Available at <http://www.euskaltzaindia.net/euskera/>]
- Irigoyen, A. 1957. Cartas de Mariano Mendigacha a D. Resurrección María de Azkue. *Euskera* 2: 119-170. [Available at <http://www.euskaltzaindia.com/euskera/>]
- Irigoyen, A. 1986. *En torno a la toponimia vasca y circumpirenaica*. Bilbao: Universidad de Deusto.
- Iturrioz, J.L. 1985. La función de -a y de -ta a la luz de la dimensión de individuación. *Euskera* 30: 175-213. [Available at <http://www.euskaltzaindia.net/euskera/>]
- Lafitte, P. 1944. *Grammaire Basque (Navarro-Labourdin littéraire)*. Donostia, Baiona: Elkar.
- Lafon, R. 1954. Le nombre dans la déclinaison basque. *Via Domitia* 1: 111-121. [Reprint in Lafon 1999: 209-217]

- Lapesa, R. 1961. Del demostrativo al artículo. *Nueva Revista de Filología Hispánica* 15: 23-44.
- Leizarraga, I. 1571. *I. Leizarragas Baskische Bücher von 1571, herausgegeben von Th. Linschmann und H. Schuchardt*. Bilbo: Euskaltzaindia. [Reprint in 1990]
- Manterola, J. 2006. -a euskal artikulua definituaren gainean zenbait ohar. In *Studies in Basque and Historical Linguistics in Memory of R. L. Trask* [Anuario del Seminario de Filología Vasca "Julio de Urquijo" XL], J. A. Lakarra and J. I. Hualde (eds), 651-676. Donostia, Bilbo: Gipuzkoako Foru Aldundia.
- Michelena, L. 1970. Nombre y verbo en la etimología vasca. *Fontes Linguae Vasconum* II: 67-93. [Reprint in Michelena 1987: 283-309]
- Michelena, L. 1971. Toponimia, léxico y gramática. *Fontes Linguae Vasconum* 9: 241-267. [Reprint in Michelena 1987: 141-167]
- Michelena, L. 1978. Miscelánea filológica vasca. *Fontes Linguae Vasconum* 29: 205-228. [Reprint in Michelena 1987: 363-385]
- Michelena, L. 1981. Lengua común y dialectos vascos. *Anuario del Seminario de Filología Vasca "Julio de Urquijo" XV*: 291-313. [Reprint in Michelena 1987: 35-55]
- Michelena, L. 1987. *Palabras y textos*. Bilbo: EHU.
- Milsark, G.L. 1977. Toward an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis* 3 (1): 1-29.
- Otsibar, K. 2003. *Lehenago Gamere*. Sü Azia. [Available at [www.suazia.com](http://www.suazia.com)]
- Pagola, R.M. (ed), 2004. *Bonaparte Ondareko eskuizkribuak. Bilduma osoaren Edizio Digitala*. Bilbo: Deustuko Unibertsitatea. [Available at <http://tesitek.com>]
- Pérez de Lazarraga, J. c. 1564. *Manuscript*. [Available at <http://lazarraga.gipuzkoakultura.net/>]
- Plank, F. and Moravcsik, E.A. 1996. The Maltese article: language-particulars and universals. *Rivista di Linguistica* 8 (1): 183-212.
- Putzu, I. and Ramat, P. 2001. Articles and quantifiers in the Mediterranean languages: a typological-diachronic analysis. In *Aspects of Typology and Universals* [Studia Typologica 1], W. Bisang (ed), 99-132. Berlin: Akademie Verlag.
- Tartas, I. 1666. *Onsa hilceco bidia*. Orthez: Jacques Rovyer. [Available at <http://klasikoak.armiarma.com/>]
- Thomason, S.G. 2007. Social and linguistic factors as predictors of contact-induced change. Paper presented at the *Symposium on Language Contact and the Dynamics of Language: Theory and Implications*, Leipzig, 10-13 May 2007.
- Trask, R.L. 1997. *The history of Basque*. London: Routledge.
- Trask, R.L. 2003. The Noun Phrase: nouns, determiners and modifiers; pronouns and names. In *A grammar of Basque*, J. I. Hualde and J. Ortiz de Urbina (eds), 113-170. Berlin, New York: Mouton de Gruyter.

<sup>1</sup> The author would like to thank the following people: the audience at the workshop on *Language Contact and Morphosyntactic Variation and Change* for their attention, Marianne Mithun and students of her *Language Contact* 2008 winter course at the UCSB for their comments, Bernard Comrie for his support. Joseba Lakarra and Céline Mounole deserve a special mention, since many of the ideas here have been (passionately) discussed with them. Finally, I am greatly indebted to two anonymous reviewers' comments, which really helped me improve this paper. This research has been possible thanks to financial aid by the Research Department of the Government of the Basque Country.

<sup>2</sup> This deliberately non-concrete synchronic definition of what we call "definite articles" goes together with the wide range of different uses displayed by elements of different languages assumed to be demonstratives and definite articles. The diachronic and language specific view we are taking in this paper will allow us to avoid these problems.

<sup>3</sup> Of course, this is a rudimentary schema, as Himmelmann himself admits (2001: 832). We can look at many studies on different languages to have an idea of the details of this grammaticalization path; good examples are Company (1991) for Spanish and Epstein's works (1993, 1994, 1995) for French.

<sup>4</sup> As Milsark remembers (1977: 5), the term of definite article "has been used for generations in the pedagogy and scholarly description of the Indo-European languages", and its synchronic formal and semantic characterization has always been done on the basis of these languages' behaviour.



---

<sup>5</sup> I will keep on glossing it as *the*, in order to make explicit once again the parallel diachronic source both morphemes, Basque *-a* and English *the*, share.

<sup>6</sup> Some researchers have noticed (Irigoyen 1986: 86) the interesting existence of a roughly 2000 year old Latin inscription found close to south Portugal where the word *Ibarra* appears. It has been taken as a proof of an early existence of the definite article by others (Iglesias 2007), since in nowadays Basque that word means ‘the valley’, analyzed as *ibarr-a* ‘valley-the’. As long as this kind of data remain so scanty and isolated, I feel more prudent not to draw big conclusions from them. Other authors with no relationship with this tradition have also suggested the possible ancient character of the Basque definite article, although for some other reasons (Putzu and Ramat 2001: 121), and in a highly uncommitted manner.

<sup>7</sup> One could doubt about the exact nature of the plural definite article; it is most usually related to the toponymy morpheme *-aga*, and said to be more recent than the singular. This is a discussion I can’t enter now.

<sup>8</sup> A reviewer points out to me that in nowadays Zuberoan, an eastern dialect, *elibat* ‘a bunch’ is used instead of *batzu*; Otsibar’s texts offer an example of this (2003). Nevertheless, we can confirm the use of *batzu* in some ancient texts of that same dialect (Tartas 1666, Egiategi 1785). Most interestingly, we may indeed make a remark that goes on the same vein as this reviewer’s doubt about the pandialectal character of *batzu*: the contiguous eastern dialect, the extinct Roncalese, has a further item besides *batzuk*, seemingly also based on the numeral *bat* ‘one’; these forms are *banak* (absolute) and *banek* (ergative), whose exact morphological nature isn’t clear for me, but significantly seem to bear D-element based plural markers. In fact, contrary to what Azkue’s dictionary says (1905-1906: 138), I couldn’t find a single instance of Roncalese *batzuk* in the texts I looked up (Irigoyen 1957, Pagola 2004). A specific study on the evolution of these eastern forms lacks, which on its turn would also shed light on the diachronic evolution of the use of *batzu*.

<sup>9</sup> I would like to thank the help Max Hofheinz and Ursula Laarmann offered me for the exact understanding of these texts.

<sup>10</sup> It always depends, of course, on what we mean with “original”. Here I am talking about (and I believe Haase wanted too to talk in the same terms) the possible (remember we have no data available) system of Basque before contact with Latin and subsequent Romance languages.

<sup>11</sup> I owe these specific data to a reviewer. I would also like to note that example (11b) may indeed be a correct one in nowadays central Basque given the appropriate context, and most surely built on the basis of a particular Spanish model construction. As a first approach to the data, I would say that the *-a* bearing phrase in (11a) wouldn’t be a referential one, but rather some sort of kind reference; its straightest Spanish counterpart would be *Eneko es buen hombre*, a not so “good” sentence to me, especially when compared to *Eneko es buen chico* ‘Eneko is a good boy’. The (11b) example would ideally stand for Spanish *Eneko es un buen hombre* (or maybe for *Eneko es un hombre bueno?*), a noun phrase with presumably a higher degree of referentiality and probably of emphatic expressiveness. The borderline between different readings is often fuzzy; the different readings of these constructions, together with the effect of model constructions’ readings on replica constructions are interesting sides of a multi-faceted discussion I cannot enter now.