A numerical review of three Basque dialects

WIM JANSEN

(European Space Agency)

1. Introduction*

An attempt is made to establish some quantitative affinity relations among three (sub)dialects of Basque: the Aezcoan variety of Western Low-Navarrese (A), the Salazarese variety of Eastern Low-Navarrese (S), and the now extinct Roncalese dialect (R). The study is based on L. Michelena's "Un vocabulario aezcoano, salacenco y roncalés preparado por el Príncipe Bonaparte" (*BAP* 1958, reprinted in *SHLV*) as the only source of information. The areas occupied by the three varieties of Basque are shown in figure 1. In this figure the Aezcoan area is delimited by the dotted line around the towns of Abaurrea Alta, Abaurrea Baja, Aria, Arive, Garayoa, Garralda, Orbaiceta, Orbara and Villanueva. The Salazarese area is taken around the towns of Escároz, Esparza, Izalzu, Jaurrieta, Ochagavía and Oronoz. The dotted line around Garde, Isaba, Roncal, Uztárroz, Urzainqui and Vidángoz defines the Roncalese dialect area. In reality the three areas are contiguous. In this context, however, they are shown as three separate dialect islands, thus allowing a better pictorial representation of the numerical results of the study.

2. Methodology

The 700 dictionary entries, collected and recorded at a time when the Roncalese dialect was still alive, are subjected to a lexical comparison (subscript: L), and to a phonetic, or sound based, comparison (subscript: PH). The ratio of lexical agreements between any pair of varieties to the total number of lexical items of the comparison is defined as the lexical affinity index between these varieties: αL. Similarly, the ratio of phonetic agreements between any pair of varieties to the total number of phonetic items of the comparison is defined as the phonetic affinity index between these varieties: αPH.
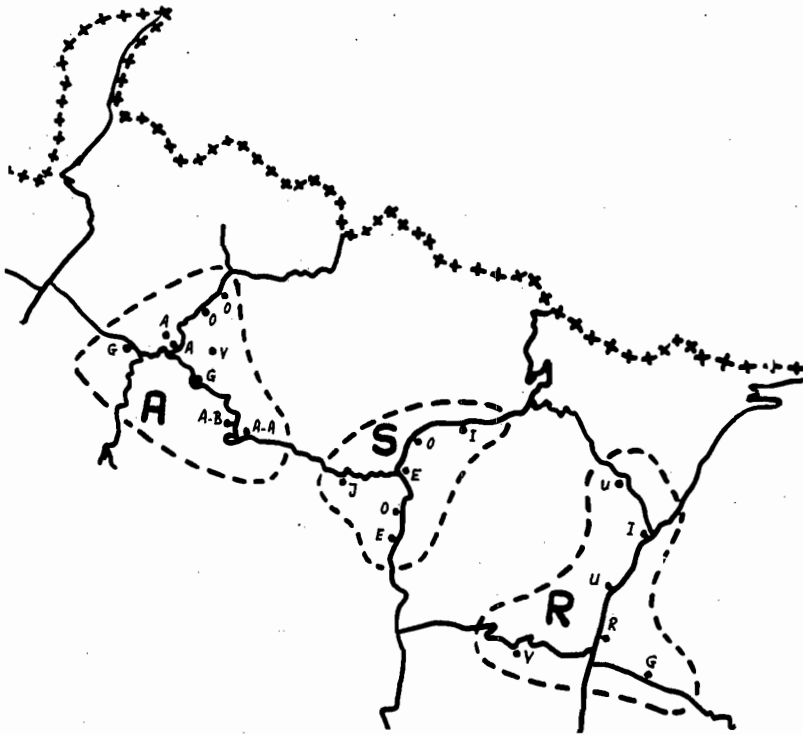
Fig. 1: The Aezcoan, Salazarese and Roncalese dialect areas

In the context of this study the two variables "lexicon" and "sound" are not in-dependent: phonetic (dis)agreements are established for pairs or triplets of items al-ready found to agree in the lexical domain. In other words, lexical situations like (AS)L can lead to phonetic situations like (AS)PH or (X)PH, but to no other combi-nations in the sound based domain. Hence, the numerical results of the statistical exercises in the lexical and the phonetic domain are expected to be very similar. The details of the αL and αPH calculations are presented in section 4 below, with an illustration of the results in figures 2 and 3.

A slightly different approach allows us to obtain a two-dimensional picture of the investigated varieties. The subset of identical lexical elements shared among A, S and R, i.e. (ASR)L, can be looked upon as a supradialectal stock, on which sound shift rules peculiar to A and/or S and/or R may have operated at later stages. The phonetic preference indices πPH indicate the percentual distribution of the lexical items from the common stock over the different phonetic subsets. These indices can be interpreted as the degree with which the different varieties hold together, or break up, from a diachronical point of view. For details reference is made again to section 4, in particular the tables 1 and 2. The subsets to the right of (ASR)L identify the synchronical lexical affinities among the three varieties at the time when the data base was collected. The πPH values derived from these subsets indicate the percen-tage of lexical items which remain phonetically stable within each pair of varieties.

3. Analyses

Although the number of entries in the reference dictionary totals 700, the summation of all identified AS, SR, AR, ASR and X subsets in a particular domain does not necessarily yield the same number 700. In actual fact, the summation of all subsets may even differ between the lexical domain and the phonetic domain. In order to explain this, a brief account is given of the calculation method, with examples from both domains.

Entry 41, F 'soir', appears as A *artsalde, arts*; S *artsalde, arts*; R *arrastri*. In the lexical domain this entry is counted *twice*, both times in the subset (AS)L, because of the identical pairs A=S *artsalde*, and A=S *arts*, both differing from R *arrastri*. Because of the sound identity between the same pairs, this double representation also applies to the other domain: 2 x (AS)PH.

Entry 105, F 'mare', appears as A *butzu*; S *putzu, butzu*; R *putzu*, suggesting a co-occurrence of two variants in S. In the lexical statistics it is represented only *once*, i.e. as (ASR)L, but in the sound based statistics *twice*, i.e. 1 x (AS)PH, and 1 x (SR)PH.

Some more examples of multiple representations of single dictionary entries are: no. 224 F 'narine', represented by 2 x (ASR)L, because of A *sudurtzilo* = S *sudurzilo* = R *sudurxilo*, together with A *sudur* = S *sudur* = R *sudur*. The former triplet is assumed to reflect three different sound variants (rendered in the spelling by *tz, z* and *x*), hence yielding 1 x (X)PH, whereas the latter triplet reflects a sound identity of all forms, so 1 x (ASR)PH. No 433, F 'oiseau', appears as A *egazti, chori*; S *abe, chori*; R *abe, chori*. The different co-occurrences suggest (ASR)L fot the overall *chori* preference; (SR)L fot the *abe* preference in S and R, contrasting with either *egazti* or *chori* in A; and (X)L covering the case of disparity among all, i.e. A *egazti*, versus S *abe*, and R *chori*. Similar solutions are then assumed in the sound domain, i.e. 1 x (ASR)PH, 1 x (SR)PH and 1 x (X)PH.

In most of these cases the total number of subsets is the same for the lexical domain and the sound-based domain, but when we look back at the dictionary entry 105, we see *two* phonetic subsets, but only *one* lexical combination.

Whenever a translation in either A, S or R is missing, the relevant entry is not counted at all. See for instance the entries 50 through 57, 115, 193, 194, and others. Whenever question marks appear in two or more columns of the dictionary, the entry is also skipped from the count, e.g. entry 288 (all three columns with a question mark), and entry 296 (A and S marked).

4. Results

One-dimensional, lexical, approach
The summation of the lexical agreements gives the following results:

$$
\begin{array}{ll}
(ASR)L & = 436 \\
(AS)L & = 98 \\
(SR)L & = 78 \\
(AR)L & = 13 \\
(X)L & = 40
\end{array}
$$

with the total number of lexical subsets 665.

In order to define the relative affinities between A, S and R, the two extremes of the scale are discarded, i.e. the situation of perfect identity ASR, and the situation of absolute disparity X. The relative lexical affinities are then calculated as follows:

$$(AS)L + (SR)L + (AR)L = 189, \text{ and}$$
$$\alpha L(AS) = 98/189 = 0.518$$
$$\alpha L(SR) = 78/189 = 0.413$$
$$\alpha L(AR) = 13/189 = 0.069$$

The calculation shows the strongest bond between the Aezcoan and Salazarese varieties, a slightly weaker one between Salazarese and Roncalese, and a very low affinity between Aezcoan and Roncalese.
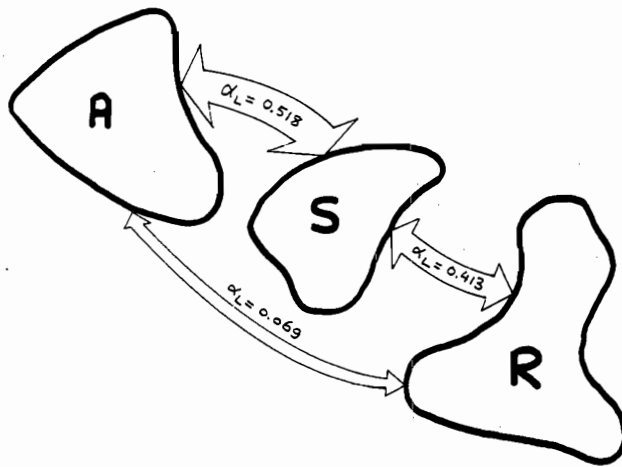

Fig. 2: Lexical affinities αL

One-dimensional, phonetic, approach
The summation of the sound-based agreements produces the following results:

$$(ASR)PH = 253$$
$$(AS)PH = 133$$
$$(SR)PH = 109$$
$$(AR)PH = 19$$
$$(X)PH = 153$$

with the total number of sound-based subsets 667.
The relative sound-based affinities are calculated as follows:

$$(AS)PH + (SR)PH + (AR)PH = 261, \text{ and}$$
$$\alpha PH(AS) = 133/261 = 0.512$$
$$\alpha PH(SR) = 109/261 = 0.415$$
$$\alpha PH(AR) = 19/261 = 0.073$$

As predicted, the relative affinities based on the phonetic criterion are very similar to the ones previously calculated in the lexical domain.
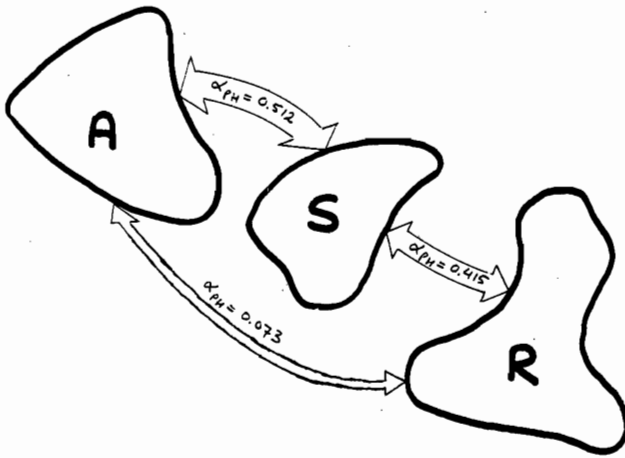
Fig. 3: Phonetic affinities αPH

Two-dimensional approach

The subset of 436 lexical items (ASR)L is taken as the corpus of a hypothetical ASR supradialectal variety, from which Aezcoan, Salazarese and Roncalese are supposed to have developed. The left-hand column of table 1 indicates in absolute numbers how the different subset derivates (ASR)PH, (AS)PH, (SR)PH, (AR)PH and (X)PH continue, or break away from, the common stock in terms of historical sound shifts. The other columns show, again in absolute terms, the phonetic stability or break-away tendency of the lexical subsets (AS)L, (SR)L, and (AR)L, which, already at the time of the data collection, were not part of the common stock.

(ASR)L	(AS)L	(SR)L	(AR)L
436	98	78	13
(ASR)PH			
254			
(AS)PH			
65	68		
(SR)PH			
56		53	
(AR)PH			
11			8
(X)PH			
50	30	25	5

Table 1

In table 2 the lexical stocks heading the four columns are normalised to 100%.
The phonetic preference indices π show the following trends:
- the common stock (ASR)L develops diachronically into AS with 14.9% pref-
  erence, into SR with 12.8%, and into AR with only 2.5%.
- the synchronical subsets (AS)L, (SR)L and (AR)L show a decreasing preference
  for stability in this particular order: 69.4% for AS, 67.9% for SR, and 61.5%
  for AR.

(ASR)L	(AS)L	(SR)L	(AR)L
100%	100%	100%	100%
πPH(ASR) = 58.3%			
πPH(AS) = 14.9%	πPH(AS) = 69.4%		
πPH(SR) = 12.8%		πPH(SR) = 67.9%	
πPH(AR) = 2.5%			πPH(AR) = 61.5%
πPH(X) = 11.5%	πPH(X) = 30.6%	πPH(X) = 32.1%	πPH(X) = 38.5%

Table 2

Both trends underline the preferential affinity between Aezcoan and Salazarese
(αL = 0.518; αPH = 0.512), followed by Salazarese (αL = 0.413; αPH = 0.415),
and a significantly weak bond between Aezcoan and Roncalese (αL = 0.069; αPH =
0.073).