

Aspiration and its historical sources: Quantifying aspiration in Zuberoan Basque

Mikhael Hayes, J. I. Hualde

Department of Linguistics | School of Literatures, Cultures & Linguistics | University of Illinois at Urbana-Champaign

Background

Since the 16th century, when the first long texts in Basque that have arrived to us were written, a striking dialectal difference that we find is the profusion in Northern texts of /h/ and aspirated stops orthographically differentiated from voiceless unaspirated ones (contrasted with the total lack of phonological aspiration in the south).

Since a phoneme /h/ (whatever its exact articulation, see Manterola & Hualde 2021, Gonzalez-Eppherre & Oihartzabal 2025) is also represented in medieval southern toponyms and in the Aquitanian-Vasconic onomastics from Roman times, it is clear that aspiration is a conservative feature in Northern Basque.

However, not all details of the distribution of aspiration in northern Basque are necessarily reconstructible for Common Basque.

Goals and data source

This study quantifies the frequency and distribution of aspiration in the Zuberoan dialect to better understand its historical sources and phonological constraints. [More broadly, this is an in-depth exploration of how generalizations about the distribution of segments may arise and weaken over time.](#)

Our data comes from Larrasquet's (1939) dictionary of northeastern Zuberoan. This source is ideal because:

- Zuberoan is highly conservative regarding aspiration.
- It shows numerous correspondences with the 16th c. texts of J. Leizarraga, a key source for historical Basque phonology.

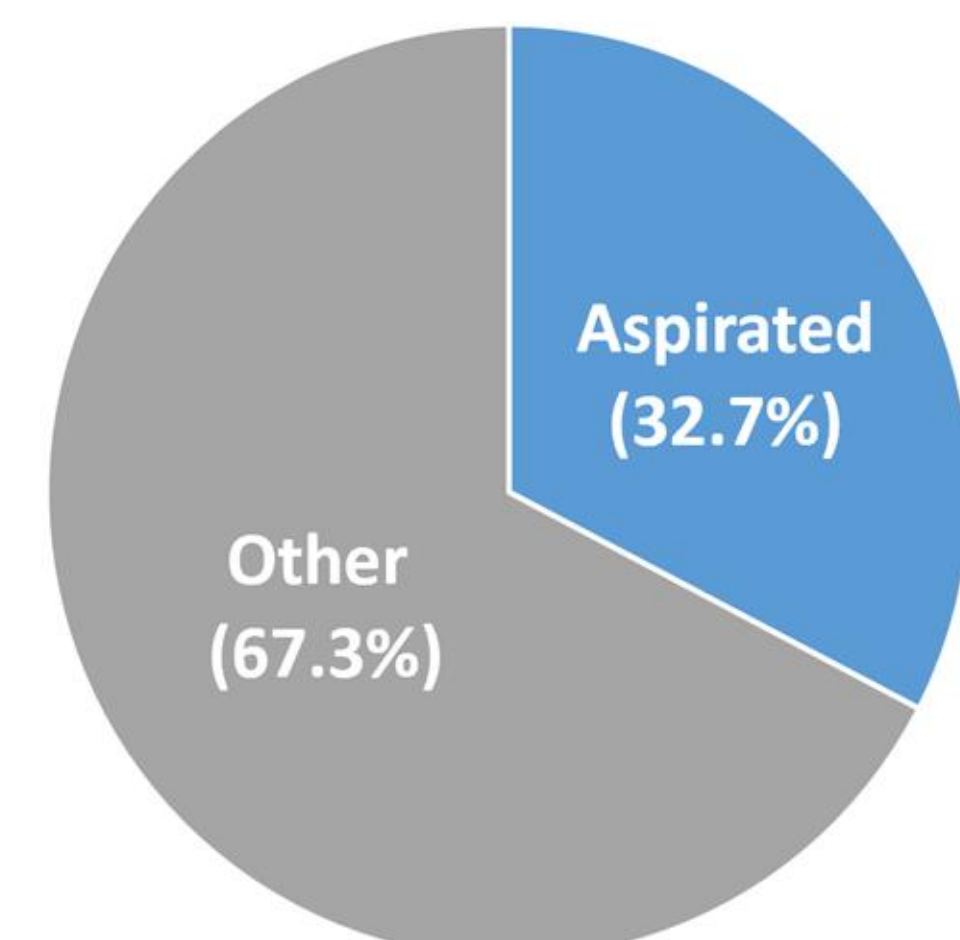
After cleaning the data (removing affixes, duplicates, etc.), our final dataset consists of 5,775 words

HATSARTU [hatsártü] vint, respirer. ⇔ Hats : hartü. ⇔ MORPH. : hatsar, inf. prés. indét. et hatsartü, respirer ; — hatsartüen et hatsartüko, devoir respirer ; — hatsartüik, respiré, ayant respiré. ⇔ Sc.

Overall frequency of aspiration

When we analyze aspiration in Larrasquet's dictionary, the first remarkable fact is that aspiration has an extraordinarily high frequency. The dictionary contains 6025 entries.

Figure 1: Aspiration frequency across entries



After discarding affixes, duplicates, finite verbal forms, interjections, etc., we obtain a total of 5775 valid entries representing words.

Of these, 1886 contain some form of aspiration. This is almost a third of all words in the dictionary (32.7%).

Sources of aspiration

- Etymological */χ/; *χarr-i
- #VnV: *anate* > *añate* (nasalized aspiration)
- Glottalization at morpheme-boundaries in compounds, later reinterpreted as aspiration: *aste+arte* > **aste?arte* > *asteharte*
- Stress-conditioned split /p t k/ vs /ph th kh/ (Mitzelena)

These sources have resulted in /h/ showing up in a variety of contexts, but with some restrictions (Lafon 1935[199], Mitxelena 1958)

Constraint I: **h...h

There may be no more than one aspirated segment per word, which Lafon (1935[1999:121]) characterized as a Grassmann's Law-like constraint. This is evidenced in alternations such as: *sarthü*, *sorthü* BUT *khentü*, *hartü*

Findings

The rule is robust for monomorphemes, but only sporadically applied in compounds. We found 44 exceptions (50 including derivations).

Figure 2: Exceptions to the **h...h constraint by type

compound type	example
number	<i>hiru-ehun</i> 'three hundred'
deictics	<i>halahüla</i> 'no matter how'
reduplicants	<i>khordoka-khordoka</i> 'staggering'
quasi-derivational	<i>ähätz-ēhi</i> 'easy to forget'
miscellaneous	<i>lepho-hāndi</i> 'goitrous'

HUNKHOI [húnkhòy, — ú nas.] a, gourmet (litt. *friend des bonnes choses de bouche*). ⇔ Hun, bon + *khoi*, friend de. ⇔ Sc.

APHURKOI [aphürkóy] n, trachée ⇔ *aphür* + *sf khoi*. ⇔ Sc.

Dissimilation and diacritics

The word *aphurkoi* '(upper part of the) trachea' shows a Grassmann's Law-like dissimilation, whereas *hunkhoi* 'gourmet (lit. *fond of good things in the mouth*)' violates the first constraint. In fact, such violations are fairly common, perhaps indicating later formations after the constraint ceased to be productive.

Hunkhoi also raises a note regarding Larrasquet's transcriptions: For typographical reasons he was not able to use more than one diacritic per character. He did, however, leave notes such as *ú nas*. With over 6000 entries, however, there is fair chance of forgetting to add a relevant note. This problematizes our question regarding the nasal source of aspiration.

Constraint II: **CVCVChV

Aspiration may not occur past the second syllable. Taking the participle suffix as example:

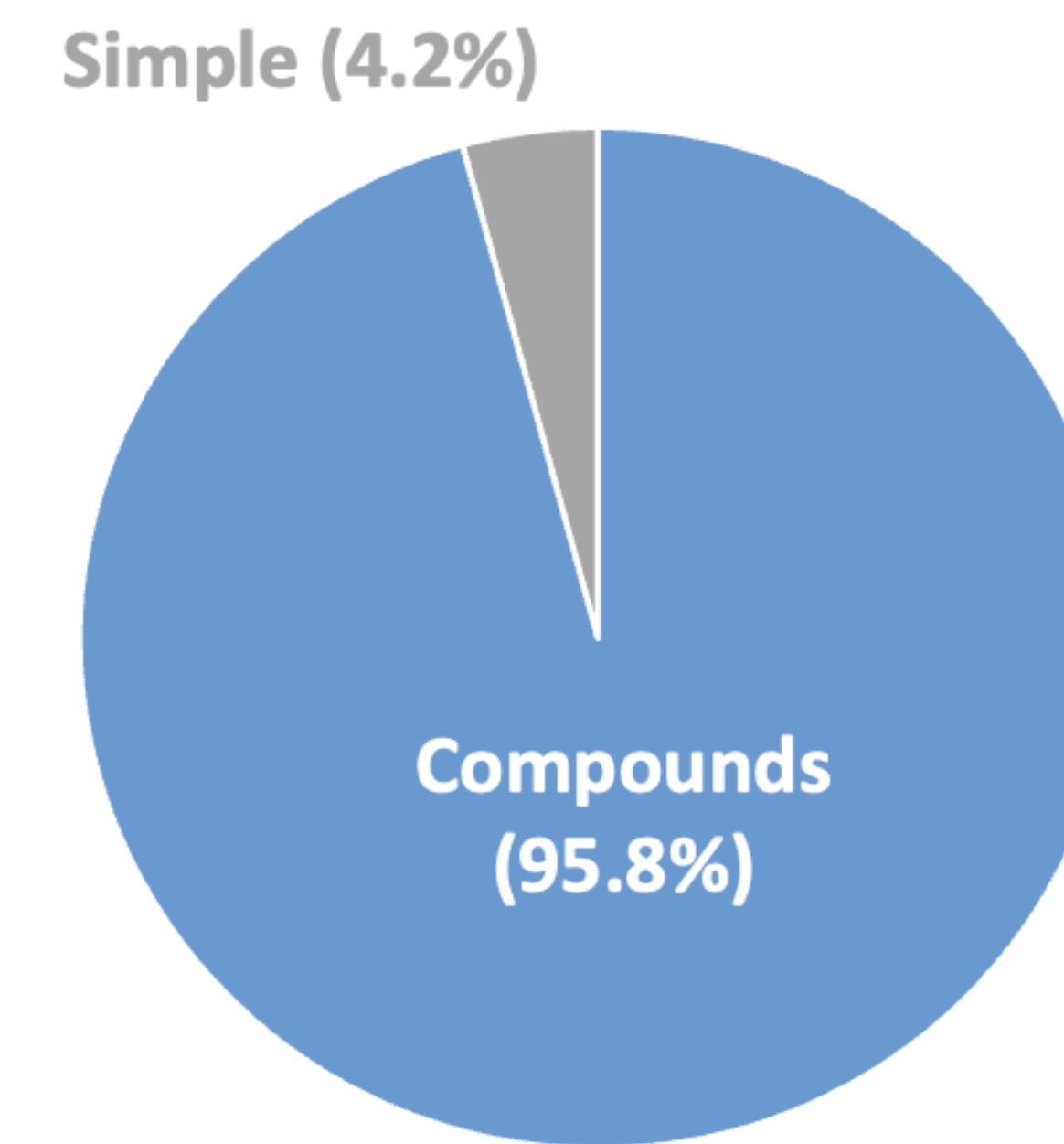
sarthü, *sorthü* BUT *agertü*

This is explained by Mitxelena (1958, 1977) as due aspiration as a reflex of stress, which in an older stage of Basque had to be on one of the first two syllables.

Findings

Aspiration is overwhelmingly restricted to the first two syllables (**#(C)V.CV.HV). However, we found 91 exceptions to this rule.

Figure 3: Proportions of late aspiration type



The majority of the exceptions to Constraint are **mostly compounds and derived words**. The derived exceptions have one of two suffixes, *-khoi* or *-ēhi*

- *-khoi*
 - *ardūkhói* (wino, *qui a le vice de boire du vin*)
 - *arakhói* (carnivore)
 - BUT *aphürkói* (trachea, lit. place where little bits get stuck)
- *-ēhi*: *entzunēhi* (easy to understand)

The compounds consist of members whose aspiration is within the first two syllables.

Figure 4: Exceptions to the **CVCVChV constraint by type

compound type	example
N+N	<i>ardi-belhar</i> 'sheep grass'
N+A	<i>jente-xēhe</i> 'common folk'
N+V	<i>merexi-ukhen</i> 'deserve'
Adv+V	<i>üntsa-ikhasi</i> 'well-mannered'
Num+Num	<i>biz-pa-hirur</i> 'threeish'

Boundary aspiration

There are a handful of compounds with aspiration at the morpheme boundary despite it being absent in the original members of the compound at that position. These violate both constraints!

- *beth-ohil* 'shifty-eyed' < *begi+ohil*
- *erth-ohāñ* 'almost free' < *erdi+dohain*
- *aste-harte* 'Tuesday' < *aste+arte*

BUT also, interesting example of dissimilation:

- *unhetsi* < **huhetsi* < *hun+etsi*

NB: boundary-induced aspiration prevails over the initial aspiration of the first member of the compound

- BUT: *hun-hartü*

Late aspiration in morphologically simple words

Just four examples, with three identified already by Mitxelena:

- *arrathü* 'rat' < Gasc.
- *baranthälla* 'February'. Perhaps a compound < *berandu+hila 'late month?' (Gonzalez-Eppherre & Oihartzabal 2025: 199)
- *borontháte* 'will' < Lat. *voluntate*
- *mārrūkha* 'wart' < Lat. *verruca* (Why oxytonic? Cf. Gasc *vorrugon*)

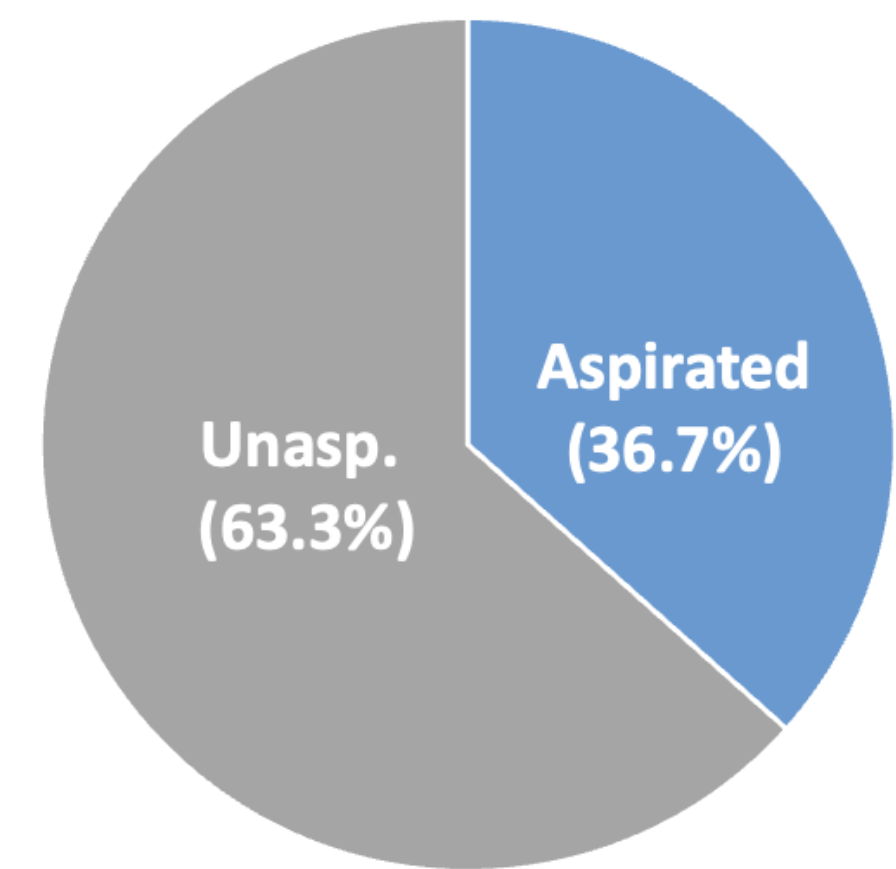
Other generalizations

Word-initial prevocalic voiceless plosives

Mitzelena suggested that words starting with aspirated /ph th kh/ may have been stressed on the first syllable at a relevant stage

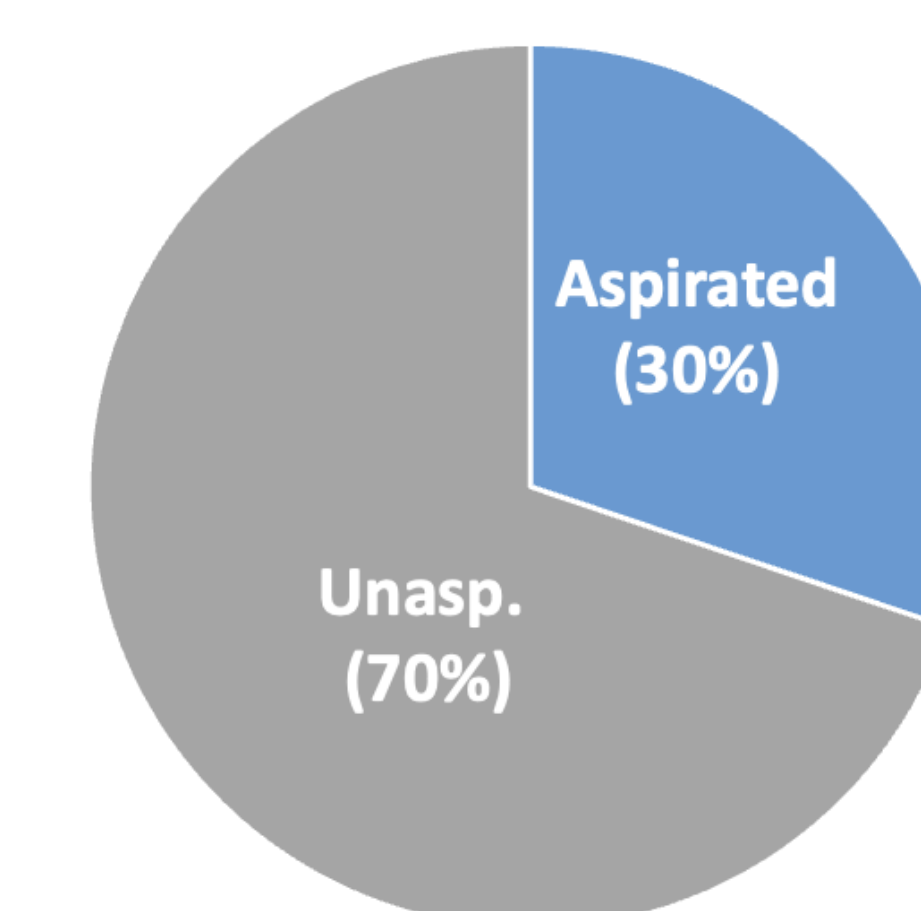
Figure 5: Initial voiceless stops

- /ph th kh/ = 138 lemmata (e.g., *khogan* 'beehive', *thumat* 'hit one's head', *pharte* 'part')
- /p t k/ = 239 lemmata (e.g., *kobesor* 'confessor', *tilur* 'linden', *partu* 'to come across/meet')



Second-syllable onset prevocalic voiceless plosives

Figure 6: [+2] voiceless stops



Mitzelena also suggested a general stress pattern of [+2] accentuation. In fact, we find aspiration only a third of the time.

aphal 'low' but *apaidu* 'food'
lakha 'grain measure' but *hoki* 'stool'
jokha 'birt' but *joka* 'jerks'

If Mitxelena was right, then many of these words must be recent to the language.

Nasalization as a source for aspiration

Zuberoan has a phonemic contrast between oral /h/ and nasal /ñ/. A key source of this nasal aspiration is the historical weakening of intervocalic *n (e.g., Latin ANATE > *ahate* 'duck').

We find that an earlier *n > h accounts for just 21 lemmate. Ex.

- **Loans**: *ühürē* 'honor' < Lat. HONORE, *xāhā* 'wash' < Lat. SANARE
- ***e-no-**: *ihür-ere* 'nobody' < *e-no-r (*ihüla* < *e-nola)
- *āhāri* 'ram' < *anari, ; *ihī* 'reed, rush' < *ini

Conclusions

1. **h...h: Systematically obeyed in monomorphemic words. Dissimilation is sporadic.
2. **h beyond the second syllable: 91 items, but only 4 unexplainable examples
3. Unexpectedly, aspirated stops are equally frequent on the onset of the first and of the second syllable.
4. The impact of -n- weakening as a source of aspiration in this position is relatively small. There are more intervocalic, non-nasal h's.