



Word-initial rhotic avoidance: a typological survey

LAURENCE LABRUNE

**Author affiliations can be found in the back matter of this article*

COLLECTION:
RHOTICS IN
PHONOLOGICAL
THEORY

RESEARCH

]u[ubiquity press

ABSTRACT

This paper addresses the issue of word-initial rhotic avoidance (WIRA) from a typological point of view. Its first aim is to document WIRA cross-linguistically, based on the examination of a sample of 200 languages designed by the WALS (Dryer and Haspelmath 2013). This set of 200 languages has been surveyed in order to reveal rhotic (and more generally liquid) phonotactic patterns in relation to word-initial avoidance. On the basis of this survey, the paper identifies two types of WIRA: i) phonological, or emic-WIRA; and ii) phonetic, or etic-WIRA. The first and most notable result of this research is that 49% of all languages containing at least one phonemic rhotic exhibit some degree of emic-WIRA, i.e, they possess no word or very few words beginning phonologically with at least one of their rhotics in their native lexicon. The paper also examines how word-initial rhotics are adapted from a non-WIRA language into a WIRA language. The loanword adaptation data suggest that WIRA is a recessive feature because no language in the sample has been observed to develop WIRA due to language contact (although one exception, Gascon, has been identified outside of the 200-language sample). Finally, the paper proposes two new universals in relation to WIRA: 1) if a language forbids /l/ word-initially, it also forbids /r/; 2) a rhotic segment never occurs as the positional allophone of a non-liquid segment word-initially.

CORRESPONDING AUTHOR:

Laurence Labrune

Bordeaux Montaigne
University, 33600 Pessac
Cedex, FR

laurence.labrune@u-bordeaux-montaigne.fr

KEYWORDS:

liquid consonants; rhotics;
word-initial position;
phonotactics; phonological
typology; universals

TO CITE THIS ARTICLE:

Labrune, Laurence. 2021.
Word-initial rhotic avoidance:
a typological survey. *Glossa:
a journal of general linguistics*
6(1): 9. 1–19. DOI: [https://doi.
org/10.5334/gjgl.922](https://doi.org/10.5334/gjgl.922)

This paper addresses a feature which has been largely overlooked in typological phonological research: the tendency of rhotic consonants not to occur in word-initial position. One very frequently comes across pithy statements such as the following in descriptive books or papers:

- (1)
- a) “r cannot occur initially in any of the Barbacoan languages that have it as a phoneme”, Curnow & Liddicoat (1998);
 - b) “/r/ does not occur initially in native Aymara”, Hardman (2001:17);
 - c) “all segmentals, except /r/, occur initially” in Haulapai (Hokan), Redden (1966);
 - d) “r never occurs word-initially” in Waskia (Trans-New Guinea), Barker (2009);
 - e) “/r/ no se registra nunca en posición inicial” in Qawasqar (Alacalufan), Clairis (1987);
 - f) “No word may begin with \underline{r} or \underline{l} ” in Hottentot/Khoisan, Greenberg (1966: 68);
 - g) “The vibrant /r/ is pronounced as a retroflexed voiced alveolar fricative word-initially. Elsewhere it is pronounced as a flap like Spanish r” in Camsá (isolate), Howard (1967);
 - h) “[r] is rare in initial position, especially so in non-loanwords” in Burushaski (isolate), Anderson (1997: 1026);
 - i) /r/ is “initially phonotactically prohibited” in Chechen (Nakh-Daghestanian), Nichols (1997: 966);
 - j) “la vibrante est absente en position initiale des lexèmes radicaux” in Susu (Niger-Congo, Mande), Houis (1963: 27).

The starting point of this research, and its working hypothesis, is that such disparate reports occur with greater than chance frequency and that they are therefore likely to reflect a general tendency of the languages of the world, which deserves systematic and general attention. The aim of this paper is to document word-initial rhotic avoidance (henceforth WIRA) cross-linguistically, with, ultimately, the more general purpose of investigating the reasons why rhotics should constitute poor word initials from a phonological perspective. This paper will focus on the first of these two issues. In section 4, it will present the results of a survey conducted upon a sample of 200 languages (based on *the World Atlas of Linguistic Structure*, Dryer & Haspelmath 2013) in order to investigate word-level rhotic distributional patterns in relation to word-initial avoidance. In addition to providing descriptive statistical data, the paper further identifies a number of representative patterns of WIRA which have been found to occur across the sample. It also considers related issues such as the role of language contact and loans in WIRA acquisition or inhibition within a language (Section 5). As a result of the close investigation of the 200-language sample, the paper proposes two novel universals (section 6). Finally, section 7 offers a conclusion and discusses a number of further issues.

But before entering the core part of this study, a number of terminological and methodological issues need to be addressed. Dealing with 200 different languages implies a large amount of heterogeneity in the data, and finding a common terminology and methodology proves to be nearly an impossible task. A number of choices have been made in order to allow for a common framework of analysis. These will be explained in Sections 2 and 3.

2 TERMINOLOGICAL ISSUES

2.1 DEFINING LIQUID AND RHOTIC CONSONANTS

The first terminological issue concerns the definition of liquid and rhotic consonants. These terms are known to be difficult to define, even though there seems to be a commonly shared intuition among phonologists over which type of segment belongs to these categories and which does not. This issue has been discussed by a number of authors (most notably Lindau 1985; Labrune 1993; 2014; 2017; Ladefoged & Maddieson 1996; Walsh Dickey 1997; Proctor 2009; Wiese 2011; etc. See also the collection of papers in this volume).

Building on these previous studies, this paper adopts a traditional and conventional definition of the term liquid. The class of liquids will be defined as including all the traditionally so-called “l”-like and “r”-like sounds. They correspond to the API symbols shown in [Table 1](#) (adapted from Walsh Dickey 1997: 11, 14, but note that lateral affricates and the uvular voiceless fricative χ have been excluded because they pattern more like obstruents than like sonorants. Note also that \varkappa is categorized as an approximant).

		DENTAL	ALVEOLAR	RETROFLEX	PALATAL	VELAR	UVULAR
laterals	Lateral approximants	l̪	l	ɭ	ʎ	ʟ	
	Lateral fricatives		ɬ ɮ				
rhotics	Trills	ʀ	r	ɽ			ʁ
	Taps or flaps	ɾ	ɽ				
	Approximants		ɹ	ɻ			ʁ
	Lateral flap		ɺ	ɻ̪			

Table 1 Liquids (= l- and r-like sounds).

If we remove all lateral segments from the [Table 1](#), we obtain the list of rhotics. In this paper, a rhotic will be defined as a non-lateral liquid segment which is proto-typically a tap, a flap, or a trill articulated in the dental, alveolar or pre-palatal region (i.e. coronal), including the retroflex place of articulation. The palatalized, velarized, glottalized, lateralized and voiceless versions of these segments (all rare in the language sample) will also be labeled as rhotics. Given this definition, any segment transcribed by means of the symbols ʀ ɾ r ɽ ɻ̪ ɻ̪̥ in the language descriptions used for this survey (as well as their counterparts involving a secondary articulation) will be automatically categorized as rhotic.

The two approximants ɹ and ɻ̪ will also be included in the class. This choice may be disputable, but note that only three languages of the sample (Diola-Fogny, English and Shipiho-Konibo) have ɹ as the sole phonological rhotic of their system, and none has ɻ̪ without having any other rhotic.

The case of the uvulars /ʀ/ and /ʁ/ deserves special attention. The uvular trill ʀ and the voiced approximant (or fricative) ʁ are reported to exist as phonemes in 12 languages of the sample. In this paper, the inclusion of these two segments in the rhotic class will be system-dependent, or conditional, which means that their categorization as rhotic or non-rhotic will depend on their status within the phonological system that hosts them and shall be established on a case-by-case examination. The general principle is that when a language has a phonemic uvular trill and/or a voiced uvular approximant in addition to one or several phonemic coronal rhotic(s), these uvulars have not been included in the sub-inventory of rhotics. This is the case of Abipon, Egyptian Arabic, Armenian, Ingush, Kayah Li and Nivkh. When a language has one or more than one uvular rhotic – /ʀ/ or /ʁ/ –, and no phonemic coronal rhotic, this/these uvular/s has/have been counted as (a) rhotic(s), except when there exists explicit and convincing evidence in the phonology of the language that /ʀ/ or /ʁ/ do not behave as sonorants but rather pattern with some other non-liquid segment. Following this criterium, the uvulars of French, German, Greenlandic (West), Hebrew and Yup'ik (Central) have been counted as rhotics. Armenian, on the other hand, is classified as having three liquid consonants, /l/, /r/ and /ɻ̪/, and two rhotics rather than three or even four (thus excluding /ʁ/ and /ɻ̪/ from the list of rhotics). This contrasts with French, which has a /ʁ/ and a /l/ but no phonemic coronal rhotic distinct from /ʁ/, so French is counted as having two liquids among which one is a rhotic (modern French /ʁ/ also has [r] as one of its allophones, and [r] is known to have been the original realization of what has become /ʁ/ – [ʁ], [ʀ], [ɻ̪], [r] etc. – in contemporary French. A similar development has occurred in German). Along the same reasoning, Lakhota is analysed as having only one liquid, /l/, because its /ʁ/, /ɻ̪/ and /ɻ̪̥/ are not considered sonorants by Lakhota specialists and they pattern with the fricatives.¹ Lakhota is thus considered as having only one liquid phoneme, /l/ and no rhotic. It should be noted that ʁ and ʀ are relatively rare as phonemes in the languages of the sample, be they considered as rhotics or not (12 languages² have one of them or the two of them) so a different choice would have had no major impact on the overall results of this study. The exaggerated importance that ʀ and ʁ have been granted in the class of rhotics undoubtedly comes from their salience in the dominant languages French and German, but from the point of view of the present research, they appear as marginal elements.

1 I am grateful to an anonymous reviewer for providing detailed information and references concerning Lakhota.

2 These 12 languages are: Apibon, French, Georgian, German, Greenlandic, Hebrew (Modern), Ingush, Lakhota, Lezgian, Nivkh, Yukaghir and Yup'ik (Central).

Many languages have some sort of r and l sounds in free variation, or use the two articulations as allophones of a single phoneme. A close examination of the phonological and phonetic structure of the 200 languages under scrutiny has revealed that when a language has only one phonemic liquid (conditional or unconditional) segment – and this is the second most common pattern found in the sample (54 languages, see annex 1) – it sometimes happens that this unique liquid is transcribed as a lateral (we will delve into the caveats of transcription below in 3.3), even though it may have a genuinely rhotic allophone as in Korean, Luvale, Meithei, Nahuatl, Sanuma, for instance. In such cases, this unique lateral liquid has been labeled as rhotic. Similarly, a rhotic phoneme noted as /r/ for instance may frequently have a lateral allophone, especially in systems which have only one alveolar liquid.

A difficult case is when one or several lateral approximant symbols represent the only phonemic liquids within their system and are not explicitly reported as contrasting with a prototypical rhotic as defined above, nor as having a rhotic allophone.³ In such cases, the language has been considered as lacking a rhotic phoneme, although the possibility that the lateral phoneme also has an unreported rhotic allophone cannot be dismissed, especially in the case of under-described languages (which is what are almost all the languages which fall into this category). 15 languages in the sample correspond to this case: Araona, Dani (Lower Grand Valley), Imonda, Kiowa, Kongo, Koromfe, Lakhota, Mandarin, Miwok, Ndyuka, Oneida, Passamaquoddy-Maliseet, Pomo (Southeastern), Supyire, Vietnamese.

This being said, one should keep in mind that the most represented liquid system in the languages in the sample (72 languages, 36%) is quite expectedly a system that combines a coronal lateral /l/ and a coronal rhotic /r/, which is phonetically almost always a dental or an alveolar tap or trill.

The set of segments which will be recognized as rhotics for the present study are given in **Table 2**. Core, or unconditional rhotics, appear in bold.

	CORE (UNCONDITIONAL) RHOTICS			SYSTEMIC (CONDITIONAL) RHOTICS
	DENTAL	ALVEOLAR	RETROFLEX	UVULAR
Trills	ɾ	r	ɽ	R
Taps or flaps	ɾ	r		
Approximants		ɹ	ɻ	ʀ
Lateral flap		ɺ	ɽ	

Table 2 Rhotics.

2.2 DEFINING THE TERM “WORD”

Another term which requires clarification is that of “word” as used for instance in the recurring key expression of this paper “word-initial rhotic avoidance”. This term should be understood in its broadest acceptance. “Word” denotes a morpho-lexical unit, whose nature and status is obviously likely to vary across languages, and it is used in the context of the present study as a cover term for lexeme, wordform, morph, stem, root or base, depending on the languages and on the theoretical stands of the linguists who have provided the description and analyses on which I have relied.

A not uncommon case in the database is when a language accepts affixes or clitics beginning with a rhotic, but not full lexical words, as, for instance, Guarani (Dooley 2006), Japanese (Labrone 2014) or Kayardild (Round 2009). In such cases, “word” will mean “full lexeme”. In a language where no morphological unit, either autonomous or non-autonomous, accepts a rhotic in the initial position, the term “word” will have a broader acceptance and will have to be

³ Here, the most common case is when a language possesses one liquid represented by the coronal lateral /l/ (18 languages). There are also 12 languages with two laterals and no genuine rhotic. These laterals are /h/, /k/ or some other type of lateral phoneme in addition to a “plain” apical /l/. Languages with three laterals and no genuine rhotics are rare (only four: Haida, Nez Perce, Squamish, Zulu), and languages with four or more laterals lacking a rhotic phoneme do not occur in the sample.

understood as “morph” or sometimes “stem”. Here again, it is impossible to achieve a perfectly satisfactory terminology, given the heterogeneity of the sources as well as specific properties of the various languages.

2.3 DEFINING THE TERM “INITIAL”

A related issue to that of “word” is what is meant by “initial position”. I take the term “initial” to mean the first phonological segment of a word, and the one which is most likely to be denoted in the orthography of the language – when the script is phonographic – or in the phonological transcriptions by linguists. But one should be cautious to note that this is not necessarily the form under which words are uttered at the phonetic level, either because the initial phonological rhotic may have a non-rhotic allophone word-initially, as in Awa Pit, for instance, or because it is preceded by a prothetic segment which is not consistently denoted in the orthography, as in Beja. This will lead us to distinguish two types of WIRA: phonological WIRA and phonetic WIRA. This distinction will be defined in Section 4 below.

Unfortunately, precise descriptions of both the phonological and phonetic nature of initial segments are not always available in the descriptions, especially in the case of under-described languages.

2.4 WHAT DOES “AVOIDANCE” MEAN ?

In this paper, the terms “avoidance”, or “restriction”, will be used in preference to other possible terms also found in the literature dealing with word-level phonotactics, such as “prohibition” or “absence”. Thus we shall speak of “word-initial avoidance” rather than of “word-initial prohibition” or “absence”. This is because, thorough progress into this research has revealed that the possibility for a rhotic segment to occur in the word-initial position cannot be easily reduced to a ‘yes’ or ‘no’ appreciation. In other words, we are dealing with a scalar phenomenon. It is thus more appropriate to consider things in terms of avoidance or restriction, rather than in categorical terms such as prohibition or absence. It is important to note that there are actually very few languages with literally zero (0, i.e. not even 1) words beginning with a rhotic in their entire lexicon, if periphery lexemes such as loanwords and onomatopoeic words are included (on the issue of loanwords and onomatopoeia, see Section 3.2 below). Another issue is that a given segment might be rare in the initial, but also in the non-initial (medial or final) position. Thus, a low number of words beginning with a rhotic should not necessarily be taken as revealing WIRA if the rhotic is also rare in other positions. An example can be found in Taba, in which /r/ beginning words are rare, but this phoneme is said altogether to be “relatively infrequent” (Bowden 1997: 57). Because there does not seem to be any asymmetry in the occurrence of /r/ within Taba words, the apparent rarity of this consonant word-initially cannot be considered as revealing WIRA, and Taba has thus not been labelled as a language with rare initial rhotics in the database. An extreme case is that of languages containing no phonemic liquid at all (there are actually seven such languages in the database, 3.5%) and hence no phonemic rhotic. For obvious reasons, these languages do not possess word-initial phonemic rhotics (although they may, in theory, possess phonetic ones) but they are not considered to be WIRA languages.

It appears that the best indicator of a possible word-level distributional restriction is the ratio between the number of word-initial rhotics and that of non-initial rhotics, or between word-initial rhotics and rhotics in all word positions (including the initial). For instance, in Maninka (Rovenchak 2011), /r/ occurs 16 times in the initial position of words in a textual corpus of 28.338 tokens, but 1564 times in all positions (including the initial).⁴ So, even though a number of 16 words beginning with /r/ cannot be considered especially low in the absolute, and that Maninka cannot be considered as a language with no word-initial rhotics, one can nevertheless assess that /r/ is rare word-initially in Maninka because of the positional asymmetry revealed by the statistical data. So “avoidance” should not be understood as corresponding to an absolute 0 but to a remarkably and presumably significant low frequency. Unfortunately, frequency data

⁴ These figures can be compared with those for /b/, for instance. /b/ occurs 1823 times in initial position and 3077 times in all positions (including the initial) so that according to Rovenchak (2011), no significant positional difference can be found for /b/ in Maninka, contrary to /r/.

of this sort are extremely rare for under-described languages, but the Maninka case provides a good illustration of how one can approach a better understanding and definition of what “avoidance” and “rarity” mean, especially in a cross-linguistic approach.⁵

3 METHODOLOGICAL ISSUES

This section provides relevant information concerning the language sample used for this study and the structure of the WIRA database (3.1), the status of special lexemes such as loanwords or mimetic words within a language’s lexicon from the perspective of WIRA (3.2), as well as the problems raised by descriptive and transcriptional heterogeneity across the sources (3.3).

3.1 THE LANGUAGE SAMPLE

The language sample used for this study is based on the set of 200 languages designed by the *World Atlas of Language structures* (Dryer & Haspelmath 2013) in order to serve as a representative sample of typological, genealogical and geographical diversity (see annex 2). For the purpose of this research, this set of 200 languages has been organized into a database using Excel. For each language, the following data have been collected: number of phonemic liquids in the overall inventory, sub-inventory of the phonemic laterals and rhotics and their phonetic realization, patterns of word-initial occurrence (i.e. whether at least one of the liquids undergoes word-initial avoidance and the pattern according to which it does, including the allophonic processes), sources and references, and other relevant information when necessary.

This database has then been surveyed in order to reveal rhotic (and more generally liquid) phonotactic patterns in relation to word-initial avoidance, the results of which are presented in Section 4.

Relevant information concerning languages which are not included in the 200-language sample has also been collected, but all the statistics and general observations are based on the 200-language sample, except when otherwise specified.

3.2 CORE LEXICON VS. PERIPHERY LEXICON

The core lexicon of a language is made up of all its native words,⁶ excluding mimetic words and proper names. Loanwords have not been taken into account in order to validate the observation whether rhotics are allowed word-initially in case a language allows rhotics in the initial position of loanwords but not of native words. However, loanwords generally prove to be very useful data in order to observe the behavior of word-initial liquids in WIRA languages, since loanwords reveal the strategies that WIRA languages develop when confronted with a word-initial rhotic in the source language, an issue that will be examined in Section 5.

5 Ideally, it would be desirable to adopt explicit numerical criteria to determine whether a language is a WIRA language or not. For instance, one could decide that a language which contains less than x% of its lexicon starting with a rhotic will be categorized as a WIRA language (x being dependent on the total number of phonemes of the language). Practically, however, this would be impossible to put into application because: i) we do not have reliable lexicon lists (dictionaries) of many of the languages of the sample; ii) we do not have data concerning the frequency of occurrence of phonemes within languages for most of the languages of the WALS set, and when we do, it is generally the case that loans, mimetics etc. are included in the sample of words retained for the frequency count; iii) when working with dictionaries, the problem is that orthography does not necessarily reflect phonology, so x would be difficult to compute; iv) most importantly, as already discussed for the Maninka case, it is not the absolute frequency in the initial which is relevant, but the ratio between initial frequency and non-initial one; v) finally, other considerations than the rough number of entries in a dictionary have to be taken into account. For instance, if a language has, say, 50 words starting with r, but that 46 of these entries contain the same prefix, then we are left with 4 r-beginning words (or 5 if we include one of the entries containing the prefix). These issues lead one to conclude that a “by hand”, case by case examination is the best – if not ideal – way to proceed, provided that the criteria are identical and that the descriptor/analyst is the same person. This is also the reason why intermediate labels such as “rare”, “rare?” and “present?” have been adopted in this study (see section 4.2). They serve as buffer categories and they actually reflect the fact that WIRA should be regarded as a scalar phenomenon rather than as a dichotomic one.

6 Or, more precisely, of all the words which are not obviously of foreign origin to the best of our knowledge. This raises the issue of the nature of the opposition between diachrony and synchrony, and the status of fossilized features or structures that may endure in a language. In many languages of the sample, rhotics seem to be accepted word-initially, but only in words that turn out, upon closer examination, to be ancient loanwords. However, native speakers are not necessarily aware of the foreign origin of these words. “Loanwordness” is actually not a unitary quality. Some words are more loanword-like than others.

Let us consider a few concrete examples. Following this criterium, Imonda, for instance, has been considered as a one-liquid, rhotic-less language, because Imonda has two liquids, a /l/ and a /r/, but the /r/, an alveolar trill, occurs only in sound imitating words (Seiler 1985).

A more delicate case is that of Drehu, in which /r/ (/or /ʀ/) is said to occur in a few loans but also in local personal names (Unë & Ujicas 1984). In the present study, Drehu has been considered as containing no /r/, but as having two liquids phonemes which are /l/ and /ʎ/.

The case of Japanese is interesting and deserves a detailed discussion because it is representative of a number of other comparable cases in the database. A superficial examination of a modern Japanese dictionary reveals that there exist thousands of /r/-beginning words in the language. At first sight, Japanese would thus appear as a non-WIRA language. However, a closer inspection shows that all these words fall into one of the following categories: loanword (mainly of Chinese and European origin), non-autonomous morpheme (suffix), mimetic word, word having undergone initial vowel deletion, or special slang word resulting from moraic inversion. In other words, it appears that /r/-initial words in Japanese are all of secondary development (Labrune 1993; 2014). In the end, only one Japanese noun which does not resort to the aforementioned origins can be found, the word *risu* ('squirrel'). Given what we know of the language, we can suspect that this word is probably a borrowing from a dialect of Chinese or from another indeterminate language. From a panchronic point of view, Japanese can thus be labelled as a WIRA language. However, this analysis is possible only because Japanese is one of the best studied languages of the world, with a long and well-documented history and a rich philological tradition. If Japanese had been an under-studied and endangered language, for which only one general grammatical description was the only available documentation, no doubt that the secondary nature of most of its word-initial rhotics would have remained ignored, and Japanese would have been excluded from the set of WIRA languages.

3.3 INVENTORIES AND TRANSCRIPTION

The main difficulty of this study lays in the lack of comparability of the sources, because the level of phonetic detail provided by different descriptors varies considerably. Transcriptions and inventories thus differ depending on authors, language varieties, the coverage of the description (whether loanwords are included or not), the theoretical approach of the author, and a number of other factors. There therefore exist true and serious transcription and comparability problems among the data. Transcriptions, especially, are more or less precise. Some sources stay at a very superficial phonemic level and do not provide all the information needed for the present research concerning allophonic variation. Such heterogeneity in the sources undoubtedly represents the main difficulty in doing research on phonological typology.

The inventories of liquid consonants used in this study all come from direct primary sources. In a second step, these primary data have also been checked in the Lapsyd database (Lyon Albuquerque Phonological Systems Database, <http://www.lapsyd.ddl.ish-lyon.cnrs.fr/lapsyd/>, referenced as Maddieson et al. 2014–2020). In a few cases, when the Lapsyd data were in contradiction with some of the methodological choices explicitly made in the present paper (for instance, if the Lapsyd recognizes as a phoneme a liquid segment which occurs only in loanwords), the sources have been checked again in order to achieve a satisfying choice meeting the criteria of the present study. A total of 30 languages of the WALS 200-language set were not yet described in the Lapsyd database (as to September 2019).

The Lapsyd inventory data has been used to double check the phonemic inventories, because, as explained on the homepage of the Lapsyd website, all the inventories provided by the database have been checked and homogenized by one unique compiler, Ian Maddieson, in an effort to provide a uniform style of analysis, particularly as it relates to the inventories of consonants and vowels. Lapsyd “selects a preferred analysis for each language and attempts to harmonize the descriptions and transcriptions across all the languages” (Maddieson et al. 2014–2020). The same type of harmonizing approach as developed by Lapsyd has been pursued for the WIRA database. To the extent that the data provided by different sources have been “filtered” by a single phonologist – the author of this paper – it is hoped that a satisfying degree of uniformity and homogeneity has been achieved.

Dictionaries have also proven a very useful source of information as far as initial occurrence is concerned, for obvious reasons. When investigating the beginning of words, it is rather easy to check how many words begin with <r> or <l> (or whatever grapheme chosen to represent rhotics) in monolingual or bilingual dictionaries which use a phonographic system of transcription, even if one should be cautious about loans, the notion of word vs. root, the use of possible prothetic sounds which are usually not transcribed in standard orthographies, and the morphological structure of the language.

Finally, when available, I have more than often relied on statements such as the ones exemplified in (1) above to decide whether a given language should be categorized as a WIRA language or not.

Due to the high number of the languages under investigation, the heterogeneity of the sources, the differences in transcription according to different authors and the lack of documentation for a number of languages, the results of this study are inevitably incomplete – if not erroneous – for a number of languages. It is certain that some occurrences of WIRA have been overlooked, and that the WIRA figures are thus under-estimated. When only one short descriptive source, and no dictionary, exists for a language, the uncertainty is especially high; one cannot be sure that a possible WIRA has not been missed, or simply not mentioned, by the descriptor, because many descriptions, especially short ones, do not provide information concerning phonotactic patterns and restrictions. However, it is hoped that this first attempt to categorize and quantify the patterns of WIRA will offer a reliable picture of a phenomenon which has been largely ignored in phonological and typological research,⁷ and that it will help stimulate further research. Any comment or complementary information on any of the languages of the sample can be sent to the author.

4 THE TYPOLOGY OF WIRA: GENERAL TENDENCIES

Let us now consider in detail the typology of WIRA in the 200 languages of the sample. WIRA is a protean and scalar phenomenon which occurs with different patterns across the languages of the sample. These various patterns of WIRA will be presented and discussed below, with examples taken from the 200-language sample. Numerical data will also be provided.

4.1 TYPES OF WIRA AND THE LEVELS OF ANALYSIS: EMIC-WIRA VS ETIC-WIRA

The level of analysis with which this study is concerned is primarily phonological. The most frequent and typical instance of WIRA in the database is phonemic in nature, and will be henceforth labeled as emic-WIRA. Emic-WIRA is represented by languages which possess one or several phonemic rhotics, yet at least one of them does not occur word-initially. In the most typical case, it means that the language has no native words beginning with at least one of its rhotic phonemes at the lexical level. This pattern is the most frequently recorded WIRA pattern in the database. It includes languages such as Basque, Japanese, Ju|'hoan,⁸ Khoekhoe, Kunama, Lak, Sango, Spanish, Turkish, Yukaghir, and many others. Note that among this category, some languages may possess several rhotics, but only a subset of the rhotics may be avoided word-initially, while others are licit, as for instance in Gooniyandi, Nunggubuya, Spanish, Trumai and many Australian languages.⁹ Further research is needed to investigate whether there exists a correlation between the number of rhotic phonemes of a given language and the manner in which a WIRA pattern occurs in that language.

⁷ Works which explicitly mention the phenomenon from a cross-linguistic point of view are Labrune (1993; 2014) Walsh Dickey (1997), and Proctor (2009).

⁸ Ju|'hoan has an unusually large number of consonants but only one liquid. It is not clear whether this liquid is phonemic or phonetic. If [r] is treated as a positional allophone of /d/ medially, Ju|'hoan should be regarded as an etic-WIRA language. If [r] is analyzed as phonemic, it becomes an emic-WIRA language, because /r/ is never found in the initial position (Snyman 1975). In the database compiled for the present study, the liquid of Ju|'hoan is regarded as phonemic (following Snyman 1975) and Ju|'hoan is thus categorized as an emic-WIRA language.

⁹ There are 38 languages (see annex 1, Table 7) which contain more than one rhotic in the database. 14 of them belong to the Australian family.

There also exists another type of WIRA which can be labeled as phonetic, or etic-WIRA. This latter type is less easily detectable and presumably less often reported in sources but it cannot be ignored. Two different sub-types of etic-WIRA have been identified. In the first sub-type, a given rhotic occurs word-initially at the phonological and lexical levels, but not at the phonetic level, because in the initial position, it is either realized as a non-rhotic or it is preceded by a prothetic vowel. For instance, in Warao, the rhotic is a flap intervocally but always a stop [d] in the initial (Romero-Figueroa 1997). So in Warao, there exist word-initial phonemic rhotics, but no phonetic ones. Another example is Wichita: in Wichita, /r/ is nasalized in the initial position (Garvin 1950). An example of a language which adds a prothetic element to a word-initial rhotic is Tiwi. In Tiwi, a language with two rhotic phonemes, /r/ does not occur word-initially, while /ɹ/ is rare in that position. According to Osborne (1974), in the few words in which /ɹ/ occurs initially, it is often preceded by a slight introductory glide. Tiwi thus appears as a language which exhibits both emic- and etic-WIRA. In Armenian, too, a language with two rhotics, a prothetic schwa is optionally inserted before the handful of /ɹ/ beginning words (Vaux 1998: 122).¹⁰

The second sub-type of etic-WIRA pertains to languages which contain a rhotic segment which stands as the non-initial allophone of some other, non-liquid phoneme, (mainly /d/ or /t/), not as the allophone of a rhotic or of a lateral phoneme. For instance, in Koromfe, a language with no phonemic rhotic, the alveolar flap [r] is an allophone of /d/ in native words, and /d/ occurs as [d] only word-initially and after a nasal stop consonant (Rennison 1997). Another example is Dani (Lower Grand Valley), which has only one liquid phoneme, /l/, and no rhotic phoneme, but a rhotic ([r]) occurs as an allophone of /t/ intervocally (van der Stap 1966). American English, which has one phonemic rhotic /ɹ/ which appears word-initially, is not an emic-WIRA language, but it is an etic-WIRA language because an alveolar flap [r] occurs as the intervocalic allophone of /t/ or /d/.

Deciding whether a language is an emic- or etic-language sometimes poses a methodological problem because in a number of cases, for instance Dani or Koromfe, the language could probably have been described as having a rhotic phoneme, say /r/, with a word-initial allophone [d] or [t], depending on the analytical choices made by the descriptor. Yet it should be noted that in languages such as Dani and Koromfe, even if the rhotic allophone had been granted the status of representing the phoneme in preference to the non-rhotic allophone, or if the language had been analyzed as containing two different phonemes standing in complementary distribution, the language would still be classified as an etic-WIRA or emic-WIRA language according to the approach¹¹ followed in this study.

There are thus, strictly speaking, two types of WIRA that need to be distinguished: phonological, or phonemic WIRA (= emic-WIRA) and phonetic WIRA (=etic-WIRA). These two major types are synthesized in [Table 3](#).

Emic-WIRA (phonological)	a) the language does possess at least one phonemic rhotic, and one of these at least does not occur in the word-initial position of words of the native lexicon
Etic-WIRA (phonetic)	b) words of the native lexicon may begin with a phonological rhotic but this rhotic undergoes mutation or prothesis at the surface level
	c) no word of the native lexicon begins with a rhotic but a rhotic occurs as a positional allophone of a non-rhotic phoneme in the non-initial position

Table 3 Types of WIRA languages.

¹⁰ Interestingly, cases of prothetic vowel insertions are often described as “optional” or “speaker dependent” in the sources, whereas other types of etic-WIRA less often are.

¹¹ Such complex cases are rare in the sample. The most delicate one is found in Khoekhoe, where a rhotic deemed phonemic by Brugman (2009) stands in complementary distribution with a non-rhotic phoneme, /t/, except at the beginning of a number of suffixes. In other analyses (Benveniste 1939; Greenberg 1966), the two are regarded as allophones of a unique, non-rhotic phoneme, because (presumably) only the root inventory is taken into account. Along Brugman’s approach, which has been adopted for this study, Khoekhoe is an emic-WIRA language, along Greenberg’s and Benveniste’s, it would be an etic-WIRA language. A similar case occurs in Bribri, which has been categorized as a one liquid/one rhotic language following Chevrier (2007), but other authors posit up to three different liquids in Bribri. Bribri is both an emic- and etic-WIRA language. See also the comment on Ju|’hoan in footnote 8.

Recall that there are also non-WIRA languages, which are of two different types, too: languages with one or several phonemic liquids which appear word-initially with no restriction, and languages which lack both phonemic and phonetic rhotics. Among the latter type there is a rather high proportion of languages for which we lack precise and detailed descriptions, especially concerning the possibility of an etic-WIRA feature.¹²

4.2 EMIC-WIRA

After defining the various degrees of emic-WIRA, this section provides the general statistics for this type of WIRA. Emic-WIRA has been categorized along a scale of six values which serve to identify and label the different word-level distributional patterns displayed by rhotic phonemes, as well as the level of information which could be gathered for each language. The status of each of the 200 languages of the sample with regards to these labels can be found in annex 2. The values are as follows:

- ABSENT: there is at least one rhotic phoneme in the language which does not occur word-initially in native words. Following the criteria adopted for this study (see Section 3), a handful of exceptions (onomatopoeia, etc.) are tolerated.

Examples: Aymara, Basque, Burushaski, Japanese, Trumai, etc.

- RARE: there exist word-initial rhotics in the native lexicon of the language, but they represent a seemingly low proportion of the lexicon.

Examples: Daga, Kera, Maricopa, Selknam, Swahili, etc.

- RARE?: there exists a number of word-initial phonemic rhotics, which seem to represent a relatively low proportion of the lexicon, but the asymmetry cannot be fully ascertained. Further research, or a better first-hand knowledge of the language could reveal that these words are loans.

Examples: Arapesh, Chinantec (Lealao) (only two languages).

- PRESENT?: word-initial phonemic rhotics seem common, but additional research should be conducted because there exists a slight suspicion that these word-initial rhotics might be limited to certain types of words (loans) or that they may be rather interpreted as reflecting an etic-WIRA (prothetic vowel or word-initial allophony not transcribed in the standard spelling).

Examples: Diola Fogny, Huitoto (Minica), Krongo, Lango, Ngiti, etc.

- PRESENT: word-initial rhotics exist in the language with normal frequency.

Examples: Arabic, Cayuvava, English, Maori, Quechua (Imbabura), etc

- IRRELEVANT: this label is used for languages which possess no phonemic rhotics in their inventory.

Examples: Ekari, Ket, Supyire, Pirahã, Usan, etc.

Note that all the languages coming under one of the above labels may also exhibit etic-WIRA in addition to emic-WIRA.

The detailed figures of emic-WIRA in the 200-language sample appear in [Table 4](#). The five first categories exclusively concern phonemic rhotics, thus providing data for emic-WIRA. See also annex 2 for the complete list of languages and their WIRA status.

¹² Following a comment by an anonymous reviewer, one could ask whether one is really dealing with “avoidance” in all the subtypes of WIRA described in Table 3. This is because while vowel prothesis or initial mutation of a word-initial phonemic rhotic can be rather straightforwardly interpreted as avoidance of a given phonotactic pattern through the use of specific repair strategies, the mere absence of any words beginning with a phonemic or phonetic rhotic as well as the asymmetrical distribution of rhotic phones that occur in subtypes a) and c), could just constitute a static, non-dynamic pattern, or even simply an accidental gap rather than a strict case of avoidance if the term avoidance is understood as implying some sort of teleonomic dimension. The examination of loanword adaptation by WIRA languages, which will be undertaken in Section 6, will bring insights to this issue, which nevertheless requires further investigation, and should be, in all events, apprehended from a broader phonological perspective, detached from the mere issue of rhotics.

LABEL	N° OF LANGUAGES	%	N° OF LANGUAGES	%
“absent”	61	30.5%	78	39%
“rare”	15	7.5%		
“rare?”	2	1%		
“present?”	14	7%	81	40.5%
“present”	67	33.5%		
“irrelevant”	41	20.5%	41	20.5%
TOTAL	200	100%	200	100%

Table 4 Emic-WIRA statistics in the 200-language sample.

The number of languages which exhibit some degree of emic-WIRA in the language sample amounts to 78 out of 200 (39%), vs. 81 (40.5%) which do not.

Furthermore, it should be noted that the remaining 81 languages are not all necessarily languages which can be considered as accepting rhotics word-initially. They also include languages about which no sufficient information on the status of rhotics word-initially was available (such languages are likely to be found in the “present?” category.). This is because, when no specific information about WIRA was found for a language, this language has been classified, by default, as a word-initial rhotic accepting language. So the number of word-initial rhotic avoiding languages might be higher than indicated.

Languages pertaining to the last category (“irrelevant”) contain no phonemic rhotic, but while they do not represent cases of emic-WIRA, they may qualify for etic-WIRA (just as emic-WIRA languages may, too). For this reason, [Table 4](#) does not tell us the whole story about WIRA. We also have to survey the language sample for specific cases of etic-WIRA, because etic-WIRA languages may or may not have phonemic rhotics. This will be done in the next section.

Clearly, the number of languages which avoid rhotics word-initially at the phonemic level is strikingly much higher than expected on a purely random basis, a feature which has been overlooked, probably owing to the fact that the most studied Indo-European languages like English, French, German or Russian do allow rhotics at the beginning of words, while in Spanish, another dominant Indo-European language, WIRA is obscured by the orthography. The first finding of this study is thus that rhotic avoidance in the initial position of words constitutes a recurring structural property in the world’s languages.

4.3. ETIC-WIRA

Let us now examine etic-WIRA. As already mentioned, etic-WIRA occurs under two different sub-types: in the first sub-type (b. in [Table 3](#)), a phonemic initial rhotic undergoes mutation and is phonetically realized as a non-rhotic segment, generally a coronal stop, or is preceded by a prothetic element, always a vowel, generally a schwa. In the second sub-type (c. in [Table 3](#)), a phonetic rhotic stands as a positional allophone of a non-rhotic phoneme in the non-initial position. In both types, and putting apart the prothetic vowel cases, one observes a complementary distribution between a non-rhotic and a rhotic in, respectively, the word-initial and the non-initial position. The languages which exhibit etic-WIRA in the sample are presented in [Table 5](#), with the phonetic details of the alternation involving the rhotic segment. Note also that etic-WIRA sometimes implies neutralization, whereby the rhotic is distinctive in medial position but neutralized with some other phoneme in word-initial position.

There are 30 languages which have been identified as etic-WIRA languages, representing 15% of the sample. Note that nine of them were also in the category of emic-WIRA. It is highly probable that there exist many other cases of etic-WIRA in the sample. This type of WIRA is probably under-estimated because the level of phonetic detail which allows its identification is not always achieved in descriptions. Moreover, dictionaries may not record the presence of a prothetic vowel in *r*-initial words. Or, on the contrary, the prothetic vowel has become fossilized and lexicalized, and it is now denoted in the orthography, which makes the language look like an emic-WIRA language, as in Yup’ik (Central).

WORD-INITIAL ALLOPHONE	WORD-INTERNAL ALLOPHONE	PHONEME	LANGUAGE	COMMENTS
[t]	[r]	/t/	Awa Pit	
[t]	[r]	/t/	Comanche	“after non-front vowels, in “laxing environments” (Wistrand-Robinson & Armagost 2012) also emic-WIRA
[t]	[r]	/t/	Dani (Lower Grand Valley)	
[t]	[r]	/t/	(American) English	
[t]	[r]	/t/	Miwok (Southern Sierra)	
[t]	[r]	/t/	Sanuma	in fast speech
[t]	[r]	/r/	(Southern) Paamese	
[d]	[ɾ] (retroflex tap)	/d/	Bribri	also emic-WIRA: Bribri has a phonemic /r/ which does not occur word-initially (emic-WIRA) but it also has a [ɾ] which is an allophone of /d/ word-medially and word-finally (etic-WIRA; Chevrier 2017)
[d]	[r] or [r̥]	/d/	Diola Fogy	
[d]	[r], [l] or [n]	/d/	Grebo	
[d]	[r]	/d/	Koromfe	
[d]	[r], [r̥] or [ɾ] (free variants)	/r/	Lavukaleve	also emic-WIRA
[d]	[r]	/r/	Otomi	also emic-WIRA
[d]	[r] or [l]	//	Sentani	also emic-WIRA
[d]	[r]	/d/	Supyire	the rhotic allophone is used in a non-accented syllable non-initially (Carlson 1994).
[d]	[r]	/d/	Tagalog	
[d]	[r]	/d/	Una	
[d]	[r]	/d/	Usan	
[d]	[r] or [l]	//	Warao	
[gr̥] (rare), [d̥] or [l]	[r]	/r/	Cayuvava	
[l] or [ʎ]	[r]	/r/	Khmer	
[l]	[r]	//	Koasati	
[l]	[r]	//	Meithei	also emic-WIRA
[l]	[r] or [r̥]	/r/	Thai	certain speakers
[n]	[r]	/r/	Wichita	
Prothetic vowel [ə] (optional)	[.ɾ]	/ɾ/	Armenian	also emic-WIRA
Prothetic vowel [i] or [ə] (most speakers)	[r]	/r/	Beja	
Prothetic vowel	[r]	/r/	Ingush	a prothetic vowel is added in front of initial Proto-Nakh *r in Ingush (Nikolayev & Starostin 1994: 93) also emic-WIRA
Prothetic vowel (“schwa onset”)	[r]	/r/	Yaqui	
Prothetic vowel (= slight introductory glide transcribed as [ə])	[.ɾ]	/d/	Tiwi	also emic-WIRA

5 WIRA IN LOANWORD ADAPTATION

In many languages, even though native words lack initial rhotics, the treatment of peripheral lexemes beginning with a rhotic, especially loanwords, deserves special attention because it allows us to observe directly how a WIRA language behaves when confronted to a word-initial rhotic. Unfortunately, very few descriptive works provide any information regarding the issue of

Table 5 Etic-WIRA patterns in the 200-language sample.

loanword adaptation, and when they do, they often remain vague or laconic. This is definitively an issue for which more systematic description is needed.¹³

From the partial documentation that I was able to gather about around 30 languages of the sample concerning loanword adaptation, it appears that two broad adaptation strategies of loanword initial rhotics occur in WIRA languages:

- WIRA is no longer enforced in loans. Rhotic initial loans are adapted with an initial rhotic in the target language, so one can talk of faithful adaptation (i.e. adaptation of a rhotic as a rhotic). Two main sub-cases occur: i) the borrowing language did possess one or several phonemic rhotics, like Acoma, Armenian, Japanese, Kannada, Korean, Mangarayi,¹⁴ Nubian (Dongolese), Rama, Turkish, etc. and now allows it or them to occur word-initially in loanwords, so the adaptation process consists merely in an extension of the phonotactic possibilities of the rhotic(s); or, ii) the borrowing language did not possess any rhotic phoneme and comes to acquire one¹⁵ in loans, in various word positions including the initial, thus expanding the number of its distinctive segments, as did for example Drehu, Koromfe, Meithei, Tagalog or Zulu, etc. The new rhotic phoneme is generally a coronal tap or trill, but not exclusively and its phonetic nature seems to depend on the language from which the loans are made (for instance, Drehu seems to have acquired a /r/ in loans from French). It is not clear whether languages may develop two or more rhotic phonemes at once (see footnote 15).
- WIRA is enforced in loans. A repair strategy of the same nature as the ones illustrated in [Table 5](#) above is applied in order to make the loan conform to the phonology of the borrowing language. Although three types of repair strategies can be expected to occur, i.e. prothesis, mutation and deletion, prothesis seems to be the most frequently observed process of initial rhotic adaptation in loans, followed by mutation. Instances of deletion have not been observed in the sample, except a restricted instance of it in Korean (see below). The prothetic segment is generally a vowel, as in Basque or Koyraboro Senni, but it can also rarely be a consonant, as in Otomi, which is said to occasionally insert a (= prothetic N before) *r*-initial words borrowed from Spanish, for instance *remedio* → *Nrremedio* (orthographic forms, Hernández-Cruz et al. 2010). When mutation occurs, the rhotic seems to be most often realized as a coronal lateral ([l]) as in Siberian Nenets¹⁶ or in former loans from Quechua into Mulayq' Aymara,¹⁷ or as a voiced stop [d]), but there are very few certain examples of this latter kind in the sample.

It is important to note that these patterns are not mutually exclusive: they can co-occur within the same language, the second strategy being adopted before the first one becomes generalized. They generally reveal different temporal strata of language contact and borrowing. This is exemplified by Basque and Korean.

¹³ One can suppose that most descriptors do not find it necessary to explicitly mention the case of loans when loans just follow the rules of the native lexemes. For instance, in an etic-WIRA language such as the ones described in Section 4.3, if loans beginning with a rhotic undergo exactly the same process as native words beginning with a rhotic, no mention will be made of the phenomenon – seen as a non-phenomenon. But no mention could also mean that the language has not borrowed many words from other languages, or that the surrounding languages are also WIRA language (a situation which would hold for Australia, where WIRA is a widely spread areal feature), or that the descriptor was not interested in loanword phonology, which seems to be a common situation when describing poorly endowed languages.

¹⁴ Mangarayi has two rhotics, /r/ and /ɾ/ (a retroflexed glide). Neither occurs initially in native words but /ɾ/ occurs word-initially in a few loanwords and personal names adapted from other areas (Merlan 1982: 186).

¹⁵ The case of a language acquiring an opposition between several rhotics in the word-initial position of loans is not documented in the database. However, this could just be a consequence of the fact that languages with two rhotics are not very common (36 languages), and, among them, languages which would be a possible source for loans and in which two distinct phonological rhotics distinctively occur in the word-initial are even less common (only 9 languages). For instance, Spanish and Gascon are two major source languages for loans into Basque, and they both have two phonemic rhotics, but one of them, the tap, does not occur word-initially, so the conditions for Basque acquiring an opposition between two rhotics word-initially in loanwords are not met. The other major contact language from which Basque is borrowing loans is French, but French has only one phonemic rhotic.

¹⁶ In Nenets, an emic-WIRA language with four liquids phonemes, a lateral /l/, a palatalized /lʲ/, an alveolar or dental trill /r/ and a palatalized trill /rʲ/, the two rhotics do not occur word-initially in native words. In loans, according to Salminen (1998), #r_ is adapted as #l_ in the Siberian dialects, and as #r_ in the European ones.

¹⁷ In Mulayq' Aymara, an emic-WIRA language with one rhotic /r/, Quechua words beginning with an /r/ used to be adapted with an /l/, but this is no longer the case and the Quechua initial rhotic is now adapted as a rhotic. Spanish initial rhotics are adapted as /r/. Initial /r/ in loans receives a sibilant realization according to Hardman (2001: 35). Interestingly, Spanish initial /d/ is also adapted as /r/.

Basque has three contrastive liquids: /l/ /r/ and /r̄/. Neither /r̄/ nor /r/ occur word-initially in native lexemes, so Basque is a WIRA language. In the course of history and of language contact, loans from surrounding languages which accept rhotics word-initially have been adapted into Basque with a prothetic vowel. For instance, *Erroma* ‘Rome’, *arrazoi* ‘reason’, *arrazista* ‘racist’, *errepublika* ‘republic’, *erlijio*¹⁸ ‘religion’ (orthographic forms) etc. Note that the prothetic vowel is not always identical. However, in very recent loans, /r/ is accepted word-initially, and modern Basque now has words such as *Ruanda*, *rap*, *ravioli*, etc. with no prothetic vowel. In such cases, the rhotic is always the trilled /r̄/, never the flap.

Another interesting case is Korean. Modern Native Korean, an emic-WIRA language, has one liquid (rhotic) phoneme with two main positional allophones [l] and [r]~[r̄]. [l] appears word-finally or before or after consonants (including itself), while [r]~[r̄] occurs between two vowels. No native autonomous lexeme begins with the liquid phoneme.¹⁹ However, in the course of its history, Korean has borrowed many words from non-WIRA languages, first from Chinese and more recently from other languages, mainly European. In contemporary South Korean, Sino-Korean morphemes undergo a /r/ → [n] / # _ process, except before /i/ and /j/ (see below), while in the contemporary North Korean variety, a spelling reform has enforced the writing of the initial liquid in Sino-Korean words, and due to a process of hypercorrection, it is now phonetically realized in this position by younger speakers, but this can be seen as the result of a relatively recent and artificial development.

So, for instance, in the southern variety of the language, the Sino-Korean morpheme /rak/ (樂) ‘pleasure’ occurs as [rak] word internally in [orak] 娛樂 ‘amusement’ and as [nak] word-initially in [nagwɔn] 樂園 ‘paradise’ (the /k/ undergoes voicing in this environment). Before /i/ and /j/ (the palatal glide), the original liquid at the beginning of Old Chinese loans has been deleted, as in the Sino-Korean morpheme /ri/ 理, ‘principle’, which is realized as [ri] in [kjori] 教理 ‘doctrine’ but as [i] in [iju] 理由 ‘cause’. However, this process of initial /n/ deletion before /i/ and /j/ is of a secondary nature, and came into effect after the 15th century. It also affected word initial /n/ before /i/ and /j/ in native Korean words (for instance, /nipʰ/ → /ipʰ/ ‘leaf’). Even the first loans from European languages used to follow these adaptation patterns (Song Nak-su 1987): for instance [namani] ‘Romania’ or [nasaro] ‘Lazarus’. However, the liquid is now accepted word-initially in recent loanwords, for instance [rɛmpʰi] ‘lamp’, [radio] ‘radio’ or [riŋʰin] ‘ricin’. It is realized as [r], [r̄] or [l] depending on the speakers or on other factors. But quite interestingly, the /r/ → [n] mutation seems still persistent in young children’s speech: the first name of the author of this paper, *Laurence*, was consistently uttered as [noransi] at the turn of this century by a young child born in 1994 in Seoul.²⁰

An interesting result that emerges from the consideration of loanword adaptation and hence, from a more diachronically oriented examination of the question of WIRA, is that many languages have evolved from a WIRA language stage to a non-WIRA stage, that is, they have come, with the course of time and under the influence of language contact, to accept word-initial rhotics. The opposite case, i.e. a language which was accepting rhotics word-initially but has come to avoid them, does not occur at all in the sample, a compelling fact in itself which can be assumed to reveal a general, quasi-universal trend of WIRA as a recessive feature. However, it is necessary to mention here Gascon (Romance), which does not belong to the 200-language sample, but stands out as a unique case. Gascon is the only language which has been identified so far as having acquired WIRA by language contact (with Basque) or by substratum effect (from the Aquitanian language, from which Basque is probably a descendant) – depending on

18 The adaptation of Latin *religio* as *erlijio* may also be interpreted as metathesis. The same adaptation process occurs in other #*rel*- beginning words adapted into Basque, for instance *erlazio* ‘relation’, *erlatibo* ‘relative’, *erloju* ‘clock’, etc. However, it should be noted that metathesis as an adaptation strategy to enforce WIRA in loans has not been found in the sample outside of these Basque examples, which could suggest that it is not metathesis which is at work here, but some other phenomena, as assumed by Egurtzegi (2011) who posits a two-step evolution process: /re-/ > /erre-/ > /er-/. I am grateful to an anonymous reviewer for providing me the Egurtzegi reference.

19 There is actually one exception: it is a metalinguistic term, the name of the <r> letter in the Hangul alphabet, *riil* which was coined after the 15th century.

20 As pointed out by a reviewer, the relationship between the two allophones of the Korean liquid could also be a matter of syllabic constituency, because the asymmetry between [r] and [l] could be reduced to an onset/coda asymmetry in most cases. More research is needed on this issue, but it is worth noting that even seen from the point of view of syllabic licensing, Korean still appears as atypical because very few cases of distributional allophony between a lateral and a rhotic governed by sub-syllabic licensing (onset vs. coda) have been found in the language sample (Garo and Warao are the only other examples I am aware of).

the theory of the origins of Basque one adopts. Although a Romance language descending from Latin,²¹ a non-WIRA language, Gascon has developed a prothetic vowel [a] in words beginning with /r/, as in *arriu* ‘river’, *arròda* ‘wheel’, *arrastèth* ‘rake’ (orthographic forms), respectively *riu*, *ròda*, *rastèl* in Occitan, from Latin *rivus*, *rota*, *rastellus*.

Putting apart Gascon, there thus exists a clear directionality with respect to WIRA: a language easily evolves from being a WIRA language to a non WIRA-language, but the opposite is extremely rare. Gascon appears to be the sole example that I could find to this date.

6 TWO UNIVERSALS

The investigation conducted over the 200 languages of the sample, as well as additional documentation over several dozen additional languages have led to the identification of the following two universals:

Universal n°1 (implicational): if a language forbids /l/ word-initially, it also forbids /r/ in the same position. The reverse is not true, i.e. no language was found in which the rhotic phoneme would be allowed word-initially but not the lateral.

Universal n°2: a rhotic never occurs as the positional allophone of a non-liquid segment word-initially, whereas a rhotic segment may occur as the positional allophone of a non-liquid segment word-medially.

No exception to these two universals have been found in the 200-language sample, nor in any of the many other languages that I have investigated.

Korean, which could *prima facie* be regarded as an apparent counter example to Universal n° 1, deserves special comments. Modern Korean, as previously mentioned, has one liquid phoneme in its inventory, with two main positional allophones: the rhotic [r] ~ [ɾ] occurs word-initially (in loans) and inter-vocally (in loans and in native words), while the lateral [l] occurs word-finally and before or after another consonant (including itself). However, the Korean case is not a counter-example to Universal n°1 because the distributional constraint bearing on liquids in Korean concerns two allophones of a single phoneme, not two distinct phonemes. Universal n°1 holds for languages which possess two distinct liquid phonemes, for example, in the most common case, a lateral and a rhotic.

This being said, it is worth noting that considered in the light of the results obtained by the present study, Korean appears as a rather atypical language from the point of view of the phonology of its liquid. From the general picture that has been gained on allophonic patterns and liquids distributional properties in the previous pages, one would rather expect the Korean liquid phoneme to use its lateral allophone word-initially rather than its rhotic one in loanwords. This is obviously not what Modern Korean does, and an internal explanation for this unexpected allophonic distribution should be sought, presumably in the history of the language. It could be that, seen in the long diachronic range, Korean is presently going through an intermediate state from a two liquids phoneme system towards a unique liquid phoneme system. Actually, a number of linguists and philologists of Korean (Lee Sung-Nyong 1955, Cho Seung-Bog 1967:203, Lee Ki-Mun 1972:70, Vovin 2020) assume that Old and, for some of them, also Middle Korean had two distinct liquid phonemes. Kim Yŏng-Čiŋ (1987) even posits three different liquid phonemes for pre-Modern Korean. The typological evidence can thus bring additional arguments to the “several liquids” hypothesis of Korean, which can in turn account for the unusual phonological behavior of the liquid segments found in Modern Korean loanwords.²²

Another possible atypical case, partly resembling Korean, is Canela-Krahô. According to Popjes & Popjes (1986), Canela-Krahô has one liquid phoneme, /l/ (a voiced alveolar lateral), with a flap allophone occurring intervocally, utterance-initially and following consonants. The source

21 Indo-European, the ancestor of Latin, was also a WIRA language, but it seems very unlikely that Gascon would have inherited the WIRA feature from Indo-European. The secondary development resulting from language contact or substratum effect is a more likely hypothesis.

22 The phonology of the Korean liquid displays many other peculiar aspects. As Kim-Renaud (1975: 66) says, “the behavior of the liquid is one of the most complicated aspects of Korean phonology”.

does not mention explicitly what the realization is word-initially when the word is not utterance initial, and whether the 'following consonant' context is tauto-syllabic or hetero-syllabic, so Canela-Krahô requires further study, but the fact that the rhotic allophone is preferred utterance initially appears as rather uncommon from a cross-linguistic point of view. However, just like Korean, Canela-Krahô is no exception to Universal n°1, because one single liquid phoneme is involved in the distribution process, not two.

7 CONCLUSION AND FURTHER ISSUES

This paper has provided a sample based, quantitative description of WIRA in the languages of the world, based on a large scale survey of 200 languages chosen for their genetic, geographical and typological representativeness (WALS, Dryer & Haspelmath 2013). The results are compelling: it has been found that 39% of the languages of the sample exhibit some degree of emic-WIRA avoidance. If the languages which lack a rhotic in their phonological system are excised from the sample, emic-WIRA languages make up 49.5% of the total. Assuming that the WALS 200-language sample reflects the diversity of the languages of the world in a balanced manner, WIRA can thus be considered as a recurrent structural property of the world's languages. At a more general level, this also means that the lower ability of rhotics to stand as phonemic or phonetic word-initials should also be definitely recognized as one of the properties that constitutes the essence of rhotics as a phonological class.

This study has also offered a methodological framework for the investigation of rhotics phonotactic characteristics, showing that WIRA comes under different sub-patterns, which need to be distinguished. Two main types occur: emic-WIRA, where a language has at least one phonemic rhotic but no word which phonologically begins with at least one of the rhotics, and etic-WIRA which comes under two forms: either the language possesses at least one phonemic rhotic which occurs word-initially at the phonological and lexical levels but undergoes mutation or prothesis, or the language has a phonetic rhotic which occurs as the allophonic realization of a non-rhotic phoneme in the non-initial position.

The examination of how initial rhotics are adapted in loanwords from a non-WIRA language into a WIRA language has also brought interesting insights, which lead to posit that WIRA is a recessive feature. This is because WIRA appears to be easily lost through language contact. There is a quasi-universal tendency for WIRA languages coming into contact with non-WIRA languages to become, in turn, non-WIRA languages in loanword adaptations; only one exception, Gascon, has been found outside of the 200-language sample.

Finally, on the basis of the results obtained through the investigation of the 200-language database, two novel universals have been proposed: 1) if a language forbids /l/ word-initially, it also forbids /r/ in the same position; 2) a rhotic never occurs as the positional allophone of a non-liquid segment word-initially.

In addition to documenting and uncovering statistical patterns of WIRA, one of the goals of the present research is also to provide an analytical grid for WIRA identification and classification, in order to facilitate cross-linguistic comparisons in forthcoming studies and also to assess the particular position of any language with respect to WIRA. More precisely, we saw that WIRA can occur under different patterns, which need to be distinguished in order to allow for a more thorough classification of the phonological and phonetic phonotactic patterns of rhotics. A lot remains to be done, obviously, but in the light of the present research, we can observe with a high degree of confidence that Korean, for instance, stands out as a typologically peculiar language as far as WIRA is concerned. The phonology of its liquid(s) could thus be now re-evaluated from the point of view of its compliance with WIRA typology. Another side-result is that WIRA cannot be taken as evidence for genetic relationship, as it has sometimes been, in order to justify the inclusion of a language in a given linguistic family. For instance, the lack of roots beginning with a liquid consonant has been repeatedly interpreted as demonstrating a supposed common origin of Korean and Japanese with Turkish, Mongolian or a number of other languages. But we now know that there is just around one chance out of two that two given languages may resemble each other with respect to word-initial rhotic occurrence, so this criterium can definitely not be used to support genetic claims, and it should be just ignored when comparing two languages.

The next main issue on the research agenda on the distributional characteristics of rhotics is that of why rhotics (or liquids) occur less frequently at the beginning of words than in other positions in so many languages. The answer to this question should be sought in a number of domains: articulatory phonetics, perceptual phonetics, functional phonology, history, evolutionary linguistics, language contact, etc. Might rhotics be difficult to produce and/or to perceive and recognize in that position? Might rhotics be phonologically “weak” segments, not suited to the initial position of words where “strong” segments are preferred? But then, should not a number of other segments such as semi-vowels be also avoided in the same context? The issue of history and of evolutionary phonology is also susceptible to bring new insights to this question, which definitely requires future research.

The results unearthed by this study also raise some new research issues that deserve further investigation. First, it would be interesting to compare, from a general cross-linguistic point of view, the behavior of rhotics with that of other segments known to undergo word-level phonotactic restrictions such as /ŋ/, /h/, /ʔ/ or retroflex consonants, but also to compare, within individual languages, the phonotactic restrictions bearing on rhotic(s) and those bearing on non-rhotic segments. A similar investigation should be conducted on laterals, which are also avoided word-initially in a number of languages, albeit to a lesser extent than rhotics. The issue of the existence of a possible hierarchy among rhotics, and more generally liquids, with respect to WIRA, is also worth of interest. Furthermore, in some languages which contain more than one rhotic, only one of these rhotics may be sensitive to WIRA constraints. Do cross-linguistic generalizations emerge? What does this tell us about the nature of rhotic consonants and about a possible hierarchy among them? Another question pertains to the precise role of language contact in the inhibition of WIRA, and what it can reveal about the phonological nature of rhotics in general. Finally, it would be necessary to conduct an investigation of the geographical distribution of WIRA. This is only a short list of the very many topics of interest concerning the phonology of WIRA in the languages of the world.

ADDITIONAL FILES

The additional files for this article can be found as follows:

- **ANNEX 1.** Statistical data concerning the structure of liquid systems in the 200-language sample. DOI: <https://doi.org/10.5334/gjgl.922.s1>
- **ANNEX 2.** The 200-language sample (extracted from Dryer & Haspelmath, 2013). By language alphabetical order, with emic and etic-WIRA status. DOI: <https://doi.org/10.5334/gjgl.922.s2>

ACKNOWLEDGEMENTS

Many people have helped me in a multitude of ways in order to develop the ideas exposed in this paper, through reading and commenting earlier drafts or oral presentations of parts of this work, and overall, by sending me useful information on one or several of the 200 languages on which this research is based. First of all, I want to thank Jean-Pierre Minaudier for pointing out to me dozens of WIRA languages, and sending me the relevant references. I also want to acknowledge the help of the many linguists who have kindly and generously shared with me their first-hand knowledge of some of the languages: Marie-Hélène Avril for Arabic and Beja, Anaid Donabedian for Armenian, Vincent Collette for Cree, Françoise Guérin for Ingush and Tchechen, Mary Pearce for Kera, Edward Vajda for Ket, Joël Miro for Gascon, Francesca Merlan for Mangarayi, Aurore Monod for Trumai, and Jean-Pierre Minaudier, again, for Fennic languages. I am also grateful to Baptiste Puyo and Georg Kaiser for facilitating my access to some of the sources, and Leah Vandever for reading two preliminary versions of this paper. I also thank the editors of this *Glossa* Issue, Adèle Jatteau and Joaquim Brandao de Carvalho for inviting me to contribute a paper on the topic of rhotics, and three anonymous reviewers who provided insightful comments. I am particularly indebted to one of them for many useful remarks which greatly helped improve the contents and the methodology of this research. All remaining errors are mine.

The author has no competing interests to declare.

AUTHOR AFFILIATION

Laurence Labrune

Dept of Language Sciences, Bordeaux Montaigne University, 33600 Pessac Cedex FR & CNRS CLLE UMR 5263, FR

REFERENCES

- Anderson, Gregory D. S. 1997. Burushaski phonology. In Alan S. Kaye (ed), *Phonologies of Asia and Africa 2*, 1021–41. Winona Lake: Eisenbrauns.
- Barker, Fay & Janet Lee. 2009. A tentative phonemic statement of Waskia. Ms. <http://www.sil.org/pacific/png/abstract.asp?id=51923>.
- Benveniste, Emile. 1939. Répartition des consonnes et phonologie du mot. *Travaux du Cercle Linguistique de Prague* 8. 27–35.
- Bowden, John. 1997. *Taba (Makian Dalam): Description of an Austronesian language from Eastern Indonesia*. University of Melbourne.
- Brugman, Johanna. 2009. *Segments, tones and distribution in Khoekhoe prosody*. Ithaca, NY: Cornell University dissertation.
- Carlson, Robert. 1994. *A grammar of Supyire*. (Vol. 14. Mouton Grammar Library). Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110883053>
- Chevrier, Natacha. 2017. *Analyse de la phonologie du Bribri (Chibcha) dans une perspective typologique: nasalité et géminée modulée*. Lyon: Université Lumière Lyon 2 dissertation.
- Cho Seung-Bog (Čo Siŋ-Bok). 1967. *A phonological study of Korean*. (Studia Uralica et Altaica Upsaliensia 2). Uppsala: Acta Universitatis Upsaliensis.
- Clairis, Christos. 1987. El qawasqar. *Lingüística fueguina. Teoría y descripción*. Valdivia : Estudios Filológicos.
- Curnow, Timothy Jowan & Anthony J. Liddicoat. 1998. The Barbacoan languages of Colombia and Ecuador. *Anthropological Linguistics* 40(3). 384–408.
- Dooley, Robert A. 2006. *Léxico Guarani, dialeto Mbyá. Cuiabá : Sociedade Internacional de Linguística*. Available <http://www.sil.org/americas/brasil/publcn/dictgram/GNDicLex.pdf>.
- Dryer, Matthew S. & Haspelmath, Martin (eds.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online <http://wals.info>, Last access on 2019-02-13.
- Egurtzegi, Ander. 2011. Euskal metatesiak: Abiaturua haien ikerketarako [Basque metatheses: Starting point for their research]. *International Journal of Basque Linguistics and Philology (ASJU)* 45, 1–79.
- Garvin, Paul L. 1950. Wichita I: Phonemics. *International Journal of American Linguistics* 16(4). 179–184. DOI: <https://doi.org/10.1086/464086>
- Greenberg, Joseph H. 1966. *Language universals*. The Hague: Mouton.
- Hardman, Martha J. 2001. *Aymara*. (Vol. 35. Lincom Studies in Native American Linguistics), München: Lincom Europa.
- Hernández-Cruz, Luis, Moisés Victoria Torquemada & Donaldo Sinclair Crawford. 2010. *Diccionario del hñāhñu (otomi) del Valle del Mezquital, estado de Hidalgo*, 2nd edition (electronical), (Serie de vocabularios y diccionarios indígenas Mariano Silva y Aceves), SIL Mexico: electronic edition. Available online <http://www.mexico.sil.org/resources/archives/51534>.
- Houis, Maurice. 1963. *Etude descriptive de la langue susu*. Dakar: Institut français d’Afrique noire.
- Howard, Linda. 1967. Camsa phonology. In Viola G. Waterhouse (ed), *Phonemic systems of Colombian languages* 14: 73–87. (Summer Institute of Linguistics Publications in Linguistics and Related Fields), Norman: Summer Institute of Linguistics of the University of Oklahoma.
- Kim, Yŏŋ-Čiŋ. 1987. Kukə i yuim e tɛ hayə [About the liquid of Korean]. In *Kan-hɛ I Pyŏŋ-Sŏn Paksə hoekap kinyəm nončhəŋ*. Seoul.
- Kim-Renaud, Young-Key. 1975. *Korean consonantal phonology*. Seoul: Tʰap Čʰulphansa.
- Labrune, Laurence. 1993. *Le statut phonologique de /r/ en japonais et en coréen*. Paris : Université of Paris 7 dissertation.
- Labrune, Laurence. 2014. The phonology of Japanese /r/: a panchronic account. *Journal of East Asian Linguistics* 23(1). 1–25. DOI: <https://doi.org/10.1007/s10831-013-9117-z>
- Labrune, Laurence. 2017. More on Japanese /r/. *Journal of East Asian Linguistics* 26(3). 301–321. DOI: <https://doi.org/10.1007/s10831-017-9157-x>
- Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world’s languages*. Cambridge: Blackwell.
- Lee Ki-Mun (I Ki-Mun). 1972. *Kukəsa kəsəl [Outline of Korean history]*. Seoul: Tʰap Čʰulphansa.

- Lee Seung-Nyong (I Siŋ-Nyŋ). 1955. Ičo čoki ii l.r-im p^hyoki munče [the problem of l.r transcription in early Middle Korean]. Reprinted in *I Siŋ-Nyŋ kukčhak sŋnčip* 2 (1988), 65–90. Seoul: Min-imsa.
- Lindau, Mona. 1985. The story of /r/. In Victoria Fromkin (ed), *Phonetic linguistics, essays in honor of Peter Ladefoged*, 157–168. Orlando: Academic Press.
- Maddieson, Ian, Sébastien Flavier, Egidio Marsico & François Pellegrino. 2014–2020. LAPSyD: Lyon-Albuquerque Phonological Systems Databases, Version 1.0. <http://www.lapsyd.ddl.ish-lyon.cnrs.fr/lapsyd/>.
- Merlan, Francesca. 1982. *Mangarayi*. London: Routledge.
- Nichols, Johanna. 1997. Chechen phonology. In Alan S. Kaye (ed), *Phonologies of Asia and Africa* 2, 941–71. Winona Lake: Eisenbrauns.
- Nikolayev, Sergei L. & Sergei A. Starostin (eds). 1994. *A North Caucasian etymological dictionary*. Moscow: Asterisk Publishers.
- Osborne, C. R. 1974. *The Tiwi language*. (Vol. 55. Australian Aboriginal Studies). Canberra: Australian Institute of Aboriginal Studies.
- Popjes, Jack & Jo Popjes. 1986. Canela-Krahô. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds), *Handbook of Amazonian languages* 1, 12899. Berlin: Mouton de Gruyter.
- Proctor, Michael Ian. 2009. *Gestural characterization of a phonological class: the liquids*. New Haven, CO: Yale University dissertation.
- Redden, James E. 1966. Walapai 1: Phonology. *International Journal of American Linguistics* 32(1). 1–16. DOI: <https://doi.org/10.1086/464875>
- Rennison, John R. 1997. *Koromfe*. (Descriptive Grammar Series). London: Routledge.
- Romero-Figueroa, Andrés. 1997. *A reference grammar of Warao*. (Vol. 6. Lincom Studies in Native American Linguistics). München: Lincom Europa.
- Round, Erich R. 2009. *Kayardild morphology, phonology and morphosyntax*. New Haven, CO: Yale University dissertation. DOI: <https://doi.org/10.5281/zenodo.829760>
- Rovenchak, Andrij. 2011. Phoneme distribution, syllabic structure, and tonal patterns in Nko texts. *Mandenkan* 47. 77–96.
- Salminen, Tapani. 1998. Nenets. In Daniel Mario Abondolo (ed), *The Uralic languages*, 516–547. London: Routledge.
- Seiler, Walter. 1985. *Imonda, a Papuan language*. (Vol. 93. Pacific Linguistics, Series B). Canberra: Australian National University.
- Snyman, J. W. 1975. *Zul'hoasi fonologie en woordeboek*. Cape Town: Balkema.
- Song, Nak-Su. 1987. Irinpa hankukŏ ii tuim pŏpč'ik yŏnku [Research on the so-called 'initial rules' in Koran]. *Hankil [Hangeul]* 197. 3–39.
- Unë, Ernest & Raymond Ujicas. 1984. *Langue drehu: Propositions d'écriture*. Nouméa : Bureau des Langues Vernaculaires, CTRDP de Nouvelle-Calédonie.
- van der Stap, P. A. M. 1966. *Outline of Dani morphology*. s'Gravenhage (The Hague): Martinus Nijhoff. DOI: <https://doi.org/10.1007/978-94-017-6361-5>
- Vaux, Bert. 1998. *The phonology of Armenian*. Oxford: Clarendon Press.
- Vovin, Alexander. 2020. Old Korean and Proto-Korean *r and *l revisited. *International Journal of Eurasian Linguistics* 2(1). 94–107. DOI: <https://doi.org/10.1163/25898833-12340025>
- Walsh Dickey, Laura. 1997. *The phonology of liquids*. Amherst, MA: University of Massachusetts dissertation.
- Wiese, Richard. 2011. The representation of rhotics. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds), *The Blackwell companion to phonology*, 711–729. Malden: Wiley-Blackwell. DOI: <https://doi.org/10.1002/9781444335262.wbctp0030>
- Wistrand-Robinson, Lila & James Armagost. 2012. *Comanche dictionary and grammar*, 2nd ed. (SIL International Publications in Linguistics 92). Dallas, TX: SIL International.

Labruene
Glossa: a journal of general linguistics
 DOI: 10.5334/gjgl.922

TO CITE THIS ARTICLE:

Labruene, Laurence. 2021. Word-initial rhotic avoidance: a typological survey. *Glossa: a journal of general linguistics* 6(1): 9. 1–19. DOI: <https://doi.org/10.5334/gjgl.922>

Submitted: 18 February 2019

Accepted: 13 October 2020

Published: 27 January 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Glossa: a journal of general linguistics is a peer-reviewed open access journal published by Ubiquity Press.