



# A Visit to the Cliffs of Jokin: A Role for Phonetizers in Second Language Pronunciation and Word Learning, with an Example from the Languages of New Caledonia

Pauline Welby, Brigitte Bigi, Antoine Corral, Fabrice Wacalie, Guillaume Wattelez

## ► To cite this version:

Pauline Welby, Brigitte Bigi, Antoine Corral, Fabrice Wacalie, Guillaume Wattelez. A Visit to the Cliffs of Jokin: A Role for Phonetizers in Second Language Pronunciation and Word Learning, with an Example from the Languages of New Caledonia. SIGUL 2023: 2nd Annual Meeting of the ELRA/ISCA Special Interest Group on Under-resourced Languages, a Satellite Workshop of Interspeech 2023, <https://sigul-2023.ilc.cnr.it/>, Aug 2023, Dublin, Ireland. hal-04381129

**HAL Id: hal-04381129**

**<https://amu.hal.science/hal-04381129>**

Submitted on 9 Jan 2024

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

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

# A Visit to the Cliffs of Jokin: A Role for Phonetizers in Second Language Pronunciation and Word Learning, with an Example from the Languages of New Caledonia

*Pauline Welby<sup>1,2</sup>, Brigitte Bigi<sup>1</sup>, Antoine Corral<sup>3</sup>, Fabrice Wacalie<sup>2</sup>, Guillaume Wattelez<sup>2</sup>*

<sup>1</sup>Aix Marseille Université, CNRS, Laboratoire Parole et Langage, France

<sup>2</sup>Université de la Nouvelle Calédonie, New Caledonia

<sup>3</sup>Université Sorbonne Nouvelle-Paris 3, INALCO, France

pauline.welby@cnrs.fr, brigitte.bigi@cnrs.fr, antoine.corral@sorbonne-nouvelle.fr,  
fabrice.wacalie@unc.nc, guillaume.wattelez@unc.nc

## Abstract

In this position paper, we argue for a role for overt phonetizers in second language learning. Phonetization or letter-to-sound conversion is often used simply as a module of text-to-speech synthesis (TTS) or to create pronunciations for dictionaries. Based on evidence of the overwhelming influence of orthographic input on second language pronunciation and word learning, we argue that on their own (or coupled with TTS), phonetizers can be effective support tools for two broad groups: 1. language learners and instructors, and 2. non-specialized users. We address the issues involved and give the example of a multilingual phonetizer under development in New Caledonia, a special status collectivity of France in the South Pacific. Encountering words and names in one of the almost 30 languages of the indigenous Kanak people of New Caledonia is an everyday experience, for example, on class lists, road signs and in news articles. Pronouncing these words is often a challenge, since each of the languages has its own phonology and its own orthography. We discuss the motivation behind the phonetizer, challenges in its development, and potential applications, many of which are common to other endangered or vulnerable and under-resourced languages.

**Index Terms:** phonetizers, multilingualism, Kanak languages, Oceanic languages, second language learning, language technology for under-resourced languages, orthography

## 1 The context

“Australian rangers discover Lifou. The three Aboriginal rangers, the ‘indigenous rangers’, Brian, Lordi, and River, were surprised to discover the biotope of the Hunētē virgin forest, in the district of Wetr.” [1, our translation]

### 1.1. New Caledonia and its linguistic diversity

In New Caledonia, we routinely encounter words and proper names in the almost 30 languages of the indigenous Kanak people, for example, on class lists, road signs or in news articles. Pronouncing these words poses a challenge. Take the excerpt quoted above from a recent news article. How do you say you say *Hunētē* and *Wetr*? How about the famous *Falaises de Jokin* (the Cliffs of Jokin), also in the district of Wetr? In Drehu, the language of the island of Lifou, these place names



Figure 1: Map of New Caledonia: La Grande Terre (the main island) and the Loyalty Islands

are pronounced [hunetē], [wet], and [dōkin], respectively, and the name of the language itself is pronounced [d̥ɛhu]. These pronunciations are counterintuitive for non-Drehu speakers; Francophone tourists often say, for example, [ʒokē] or [ʒokin].

La Nouvelle Calédonie or New Caledonia, an archipelago in the South Pacific (see Figure 1), is currently a special status collectivity (*collectivité sui generis*) of France. It was the explorer James Cook, who in 1774, struck by their hills and mountains, named these islands for Scotland. Regular contact with Europeans started in the first half of the 19<sup>th</sup> century, with whalers, sandalwood traders, and adventurers, followed by missionaries. English-speaking Protestant missionaries from the London Missionary Society and French-speaking Catholic missionaries from the Marist order of priests introduced writing systems for many of the languages to further their mission of teaching the Gospel and converting the indigenous peoples to Christianity. On September 24, 1853, France “took possession” of most of present-day New Caledonia, claiming it as a colony. Today French is the predominate language of instruction in New Caledonia, in primary and secondary school, as well as at university. The 1998 Nouméa Agreement specifically acknowledges that “[t]he Kanak languages are, along with French, languages of instruction and culture in New Caledonia” (our translation; for discussion and context, see [2]). Four Kanak languages can be studied as a subject for the baccalauréat exam at the end of secondary school: A’ijë, Drehu, Nengone, and Paicî. Students majoring in Oceanic languages and cultures at the University of New Caledonia study at least two Kanak languages. Like many Pacific islands and countries, particularly in Melanesia and Australia, New Caledonia is characterized by

the diversity of its languages [3], [4]. According to recent counts, there are just under 30 Kanak languages, most of which are Southern Oceanic: 27 Melanesian languages, one Polynesian outlier language (Fagauvea), and a creole (Tayo) [5, p. 10]. Most of these languages are endangered or vulnerable, with communities of speakers ranging from fewer than 20 for Sîshëë to around 16,000 for Drehu, according to the last census [6]. In addition, many other languages are spoken, by people from (or with roots in) the Pacific, Asia, and other regions.

## 1.2. The challenges

While linguistic diversity, in New Caledonia and elsewhere, is certainly to be celebrated, it poses practical challenges.

### 1.2.1. In New Caledonia

Each of the Kanak languages has its own phonology as well as its own orthographic system. The languages have a wide range of vowel and consonant inventories, and some have lexical tone (e.g. Kwényî and Paicî). Learning the grapheme-phoneme correspondences (GPCs) of a second language (L2) is not always straightforward. As we have seen, it can be difficult to know even approximately how to pronounce words and names in the languages we encounter, much less to master a new spelling system. It is all the more difficult to simultaneously master the spelling systems of two or three — or potentially many more — languages. There is also likely to be interference from the GPCs of the L1 or those of the language in which one typically reads and writes.

To take one example of a single letter, in just a few languages [7, 8, 9, 10, 11]. In Drehu, the letter ‘r’ represents an apical trill or tap [r] or a lateral [l] (e.g. *rais* [rais] or [lais] ‘rice’). The digraphs <tr> and <dr> represent affricates (*atr* [atʃ] ‘human being; twenty’, *Drehu* [dʒehu]). In Nengone, the language of the island of Maré, ‘r’ represents [ɹ], familiar from English, (e.g. *era* [ɛɹa] ‘to sing’). In Xârâcùù, spoken in the region of Canala, Thio, and Boulouparis on the *Grande Terre*, ‘r’ between nasalized vowels represents a nasalized rhotic tap [ɾ̃] (*xârâcùù* [xârâci:]), perceptually and acoustically similar to [n], as evidenced by Pastor Maurice Leenhardt’s name for this language: *anesù* [10, p. 14]. In French, the shared language of virtually all Caledonians, ‘r’ is often a fricative [ʁ] or a trill (e.g. *caramel* [kaʁamel]), but never a nasal. Some other examples of GPC conflicts are given in Table 1.

Table 1: Examples of conflicts in grapheme-phoneme correspondences in several Kanak languages and French

| Language | GPC                     | Word  | IPA                | gloss     |
|----------|-------------------------|-------|--------------------|-----------|
| Drehu    | <ë> ~ /ɛ/<br>or /æ/     | xën   | /xɛn/              | eat       |
| Paicî    | <ë> ~ /ʌ/               | këkë  | /kʌkʌ/             | bird      |
| Kwényî   | <ë> ~ /ø/               | xë    | /xø/               | octopus   |
| French   | <ë> ~ /e/<br>or /ɛ/     | canoë | /kanoe/            | canoe     |
| Drehu    | <j> ~ /ð/               | jol   | /ðol/              | difficult |
| Kwényî   | <j> ~ / <sup>m</sup> ɟ/ | je    | / <sup>m</sup> ɟə/ | water     |
| Nengone  | <j> ~ /dz/<br>or /ɟ/    | jo    | /dzo/              | suffer    |
| French   | <j> ~ /ʒ/               | je    | /ʒə/               | I         |

### 1.2.2. Beyond New Caledonia

The motivation behind our project will resonant with other multicultural communities. The challenges addressed are not limited to New Caledonia or to the Kanak languages. They are experienced by other communities where several languages are present to one degree or another (e.g. countries with migrant communities) or in which the L1 (or the literacy dominant language) and the L2 (or L2s) have very different orthographies, for example, English and Irish [12, 13].

## 2 A role for stand-alone phonetizers

We argue here that a stand-alone, multilingual phonetizer can be an effective tool to partially address these challenges, in the Caledonian context and beyond. The proposal builds on two independent lines of research, the first on the development of technologies for under-resourced languages, and the second on the influence of orthographic input on L2 learning.

### 2.1 Phonetizers as a tool

A phonetizer is a digital tool that takes as input a word or text in standard spelling and converts it to a phonetic transcription or a pronunciation spelling. Phonetization is most often used as component rather than a stand-alone tool, for example, as a module for text-to-speech (TTS) synthesis, as a tool to create pronunciations for dictionaries, or as a step in forced alignment.

Stand-alone phonetizers are publicly available as pronunciation aids for some languages including English and French and the endangered language Occitan (see § 6). None, however, are available for the Kanak languages or for other widely used languages of New Caledonia, such as Wallisian, a Polynesian language that likely has more speakers in New Caledonia than on the island of Wallis itself.

### 2.2 Orthographic input and L2 learning: an overwhelming influence

The potential for phonetizers as L2 learning tools is grounded in research demonstrating the overwhelming influence of orthographic input on L2 pronunciation and word learning. Learning new words with their orthographic form not only influences the formation of phonological representations, as reflected in pronunciation, but can also facilitate the addition of these words to the L2 lexicon [14]. As observed by [15], “when faced with speech, which is inherently [] highly variable and fleeting, the orthographic form offers L2 speaker-listeners something stable to ‘grab on to’.”

Seeing (and so reading) the written form of a word can draw L2 learners’ attention to otherwise unnoticed aspects of the segmental structure of auditorily presented words. For example, in French obstruent+rhotic clusters (e.g. *trèfle* [tʁɛflə] ‘shamrock’), L1 Mandarin speakers can mistakenly perceive the fricative rhotic [ʁ] as aspiration. The written form of the word helps these learners perceive the presence of a rhotic [16].

Conflicts in L1 and L2 GPCs, however, can pull L2 speakers towards non-native phonological representations [14, 15]. For example, while L2 English pseudowords like *mog* [mag] were memorized and recalled more accurately by L1 French speakers when presented with both audio and orthographic input than with audio only input, they had pronunciations that were more French-like (i.e. more non-native). For example, *mog* was pronounced more like [mɔg] than [mag] (French <o> ~ /ɔ/).

These phenomena are clearly relevant to the Caledonian context, as illustrated by French-like pronunciations of Kanak place names like *Jokin* (§ 1.1). We expect that the transcriptions of the phonetizer will mitigate the influence of these L1/L2 GPC conflicts, while retaining the benefits of written input.

## 2.3 The end users and their requirements

### 2.3.1 Two broad groups of end users

We argue that, on their own, phonetizers can be effective support tools for two broad groups of end users: 1. language learners and teachers (as suggested in [12, ch. 8]) and 2. non-specialized users. The first group are people with an interest in a particular language who need to learn the correspondences between the letters and sounds of that language. This group includes language teachers and L2 learners, including primary and secondary school students, university students, and heritage speakers. The second group of users are those who occasionally or even regularly need to pronounce a word, phrase, personal name or place name in any one of the local languages. This group includes journalists, politicians, culture and tourism professionals, and teachers of all subjects. Indeed, it includes anyone wishing to mark their respect for a person and a community by making the effort to accurately pronouncing a name or a word from their language.

### 2.3.2 Requirements of the end users

The requirements of these two broad groups of end users differ in important ways. Language learners need to master the orthography of the language studied and to accurately pronounce the sounds of the language, ideally approaching the pronunciation of native speakers. For example, they need to preserve contrastive vowel length (e.g. Drehu *lolo* [lolo] ‘beautiful’ vs. *loloo* [lolo:] ‘to hurry’). Non-specialized users, on the other hand, arguably need simply to be able to pronounce a given word or name in a way that avoids a radical departure from the segment string, for example, by pronouncing *Drehu* as [dʒehu] rather than [dʒey], as if it were a French word. The output must therefore be tailored to the needs of each group. For *drehu*, a pronunciation spelling like *djé-hou* (with a reminder that the ‘h’ is not silent, unlike in French) would be more helpful for non-specialized users than an IPA transcription.

## 3 A multilingual phonetizer for Caledonians

We have recently undertaken the implementation of the first version of a multilingual phonetizer as a pronunciation and language learning tool for the languages of New Caledonia. The project is structured around four work packages.

**Work package 1.** Development of an ergonomic, easy-to-use web interface that accepts input (standard spelling) and displays output (transcriptions).

**Work package 2.** Implementation of grapheme-phoneme conversion rules, for each language. For this first version, we will start with Drehu and Paici, two of the four languages that can be studied for the *baccalauréat* (see § 1.1). Where necessary, the phonetization will include syllabification, since it is not always clear what the graphemes are, even for languages with transparent, one-to-one grapheme-phoneme correspondences. For example, French speakers are tempted to pronounce sequences of vowel letters as French multi-letter graphemes, decoding a sequence like ‘oi’ not as [o.i] (as in

Nengone *roi* [ɔ.i] ‘bon, bien’), but as [wa] (as in French *roi* [ʁwa] ‘king’).

**Work package 3.** Design of a user-friendly phonetic transcription system. The output will include two choices of transcription: 1. International Phonetic Alphabet (IPA), for the phonetically inclined, and 2. pronunciation spellings based on French orthography, enriched with pedagogical tips (e.g. illustrations, brief explanations, and sound files). This second transcription system is necessary since IPA is (alas!) not widely interpretable by the general public.

**Work package 4.** Launch of the phonetizer online, beta testing.

## 4 Approach

### 4.1 Consultation and dialogue

Native language consultants will be invaluable partners in the project. Throughout the project we will consult with native speakers from the different communities, for example, representatives from the Académie des Langues Kanak (ALK) and customary chiefs (*coutumiers*). We have begun soliciting input from the potential users (see § 2.3.1). Ongoing dialogue is essential for building a tool that is both usable and used, and that is respectful of the languages and their speakers and of the scholars on whose research this contribution is based.

In these exchanges, we emphasize that the phonetizer is complementary to the writing systems developed by generations of native speakers, missionaries, and linguists, most recently in the ALK collection *Propositions d’écriture* ‘Proposals of writing systems’ [5, 7, 9, 10, 17, 18, 19, 20, 21], and is not intended to replace these systems. To some, using a transcription system based on French may seem, at least initially, like a step backwards, reminiscent of the linguist Jean-Claude Rivierre’s critiques of the “orthographic Gallicisms and notational approximations” of the 19<sup>th</sup> century Marist missionary Fr. Xavier Montrouzier [22, p. 14, our translation]. *Au contraire*, the phonetizer is complementary to the writing systems and is intended as tool to help learners master a language, including its spelling system.

### 4.2 An incremental, open-source approach

The multilingual phonetizer will be extensible, to allow other languages, both Kanak languages and languages of other communities, to be added to subsequent versions, and to allow the transcription systems to be updated.

To facilitate the extension of the approach developed here to projects on other languages and other communities, we will adopt best practices in open science, including making our scripts publicly available (e.g. on <https://osf.io>). We ourselves will, of course, use and re-use existing code available under open-source licenses. Reuse and sharing of resources are particularly important for work on endangered and vulnerable languages. We will also make our research papers available by publishing in open access journals or depositing in open access repositories. In addition, we will make French-language versions available, an essential step in making our research accessible to the people most directly concerned by the project.

### 4.3 Implementation of the phonetizer

Converting from written text to sound symbols poses a number of challenges for any language due to: 1. the degree of lack of correspondence between the spelling of the lexical items and

their sounds, and 2. unknown words like proper names or code-switching. The automatization of phonetic transcription can be implemented in many ways, often roughly classified as dictionary-based or rule-based strategies, although many intermediate solutions exist (e.g., a dictionary-based system supplemented by a rule-based algorithm to phonetize unknown words, as in [23]). Phonetization is a structure prediction task; both the input and output are structured, consisting of sequences of letters and phones, respectively. Dictionary-based systems store a maximum of phonological knowledge in a lexicon, while rule-based systems consist of rules that are based on inference approaches from the machine learning domain or are proposed by linguists. In the project, we adopt a knowledge-based approach, since data-driven approaches require large sets of training data, and such data are not available for the Kanak languages, as is often the case for under-resourced languages.

The Kanak languages and other Oceanic languages have relatively transparent orthographic systems, with generally one-to-one GPCs, as well as predictable context-sensitive phonological processes (e.g., the nasalization of rhotics in Paicî and Xârâcùù). Exceptions lists will be compiled to take into account cases like English loan words to Drehu, which may be written with single-letter vowel graphemes, although they have long vowels (e.g. *hos* [ho:s] < *horse*). Other cases will not be so easily dealt with and will not be accounted for in this first version. These include, for example, the distinction in Drehu between some orthographically identical minimal pairs of content words and grammatical particles (e.g. *tha* ‘catch (in fishing)’ [θa] vs. *tha* (negation) [θa:], [7]).

## 5 Discussion

### 5.1 Applications

We present here three of the many potential uses of phonetizers, in the Caledonian context and beyond.

#### 5.1.1 Overt phonetization coupled with TTS

In the future, the phonetization under development here could be used in a TTS system for one or more Kanak languages, if the communities so desire. We would argue that overt phonetization, with carefully designed spelling pronunciations displayed on the screen rather than running unseen in the background, could be coupled with TTS to improve learning outcomes. This general prediction follows from results on the overwhelming influence of written input on L2 learning (§ 2.2).

#### 5.1.2 Pedagogical materials

Like many endangered and under-resourced languages, pedagogical materials for the Kanak languages are lacking. For these languages, pedagogical materials take a variety of forms (see [25, 26]). The phonetizer will be a tool for creating resources and giving new life to existing ones. For example, a richly illustrated children’s book gives the words for hundreds of objects and concepts in five Kanak languages [27], as in (1).

- |     |                       |            |
|-----|-----------------------|------------|
| (1) | Arc-en-ciel ‘Rainbow’ |            |
|     | Fwāi (Fwâi):          | pwalalu    |
|     | Yuaga (Yuanga):       | thucaabele |
|     | Paicî:                | ubê        |
|     | Cuauru (Numèè):       | tu miê     |
|     | Drehu:                | lewen      |

The multilingual phonetizer could be used to enrich these lexical entries, allowing teachers to incorporate the words in language awareness (*éveil aux langues*) activities [28, 29].

#### 5.1.3 Religious ceremonies

The phonetizer could be used as a tool to prepare Bible readings or prayers *en langue*, to be used in church services, coming back full circle to the missionaries’ goal in establishing orthographic conventions. This could be particularly useful in the many communities where most members are not fluent speakers.

### 5.2 Challenges

We are faced with a number of challenges, many of which are shared by those working on other languages. These include: choice of dialect; personal preferences in the pronunciation of given and family names; inadequate linguistic resources for some languages and a lack of human resources to remedy this problem; non-standard use of phonetic symbols in existing materials; outdated descriptions; and the handling of lexical tone. Some of these challenges may give rise to opportunities, for example, by inspiring work on the sound structure of the contemporary languages. Another challenge comes from relatively limited access to digital resources (e.g. computers and reliable, affordable Internet) in rural areas, including on tribal lands (see [24]), affecting large swaths of potential users.

### 5.3 Subsequent versions

Natural candidates for the extension of the phonetizer are the other languages with an ALK *Proposition d’écriture*, currently A’ijë, Drubea, Iaai, Kwényi, Nengone, Numèè, and Xârâcùù. Each *proposition* provides an exhaustive description of the writing system and was constructed with input from members of the language community, lending weight (if not always consensus) to the proposed conventions.

When the orthographic system of a language is transparent, with mostly one-to-one GPCs, rule-based and dictionary-based approaches produce the same result. From a computational point of view, however, it is less costly to create a pronunciation dictionary from rules and to then use the dictionary. This approach is also more collaborative since the dictionary can be made available online, and members of the language community can help to correct it. Such a crowdsourcing, citizen science approach may be considered for future versions.

In this first version, the user will manually select the input language. Automatic language detection could be integrated in future versions, using natural language processing techniques.

Although the first version will be implemented as a web interface, a smartphone app would likely be more usable/used.

## 6 Acknowledgments

We gratefully acknowledge the generations of native speakers, missionaries, and linguists who developed the writing systems upon which this project depends. We also thank the Académie des Langues Kanak, whose *Propositions d’écriture* series is an invaluable resource. We stand on the shoulders of giants.

We thank the Laboratoire Parole et Langage for providing seed funding for the project through a *bonus qualité recherche* grant.

Pauline Welby would like to dedicate this modest contribution to the late Chris Cieri (1963–2023). I will never forget Chris, his kindness, his passion for language resources, and his respect for languages and their communities. I will honor his legacy.

## 7 References

- [1] N. Esturgie. “Des rangers australiens à la découverte de Lifou”, Nouvelle Calédonie la 1ère, May 25, 2023. <https://la1ere.francetvinfo.fr/nouvellecaledonie/province-iles/lifou/des-rangers-australiens-a-la-decouverte-de-lifou-1399134.html>
- [2] Agence de Développement de la Culture Kanak, *Actes du colloque Langues kanak et l'Accord de Nouméa*. Nouméa, New Caledonia, July 2000.
- [3] G. F. Simons and C. D. Fennig, (eds). *Ethnologue: Languages of the World (20<sup>th</sup> ed.)*, Dallas: SIL International, 2017.
- [4] D. Tyron, “Language Distribution in the Pacific,” In *Oceanic Encounters: Exchange, Desire, Violence*, M. Jolly, S. Tcherkézoff, and D. Tryon (eds). Acton: Australia National University Press, 2009.
- [5] Académie des Langues Kanak, *Propositions d'écriture du iaai, langue parlée à Ouvéa, Nouvelle-Calédonie. Hna setr hwen iaai ae thep ünyi*. Nouméa: Académie des Langues Kanak, 2020.
- [6] Insee (Institut national de la statistique et des études économiques). Recensement de la population, 2014.
- [7] L. D. Sam. (dir.), *Aqane cinyihanyi gene drehu : Propositions d'écriture de la langue drehu*. Nouméa: Académie des Langues Kanak, 2009.
- [8] P. Welby, J. Monnin, C. Torres, and F. Wacalie. “Drehu diphthongs and other recent findings,” Paper presented at the 18th Australasian International Conference on Speech Science and Technology, Canberra, Australia, 2022.
- [9] Académie des Langues Kanak, *Propositions d'écriture du nâa nengone. Ka xiwaimomo pene nengone (langue parlée à Maré)*. Nouméa: Académie des Langues Kanak, 2015.
- [10] Académie des Langues Kanak, *Propositions d'écriture du nâa xârâciù. Kèè-su ngê nâa xârâciù*. Nouméa: Académie des Langues Kanak, 2014.
- [11] Nengone : Initiation à la langue de l'Île de Maré. Website. <http://nengone.univ-nc.nc/>
- [12] B. Ó Raghallaigh, *Multi-Dialect Phonetisation for Irish Text-to-Speech Synthesis: A Modular Approach*. Doctoral thesis. Trinity College Dublin, 2010.
- [13] N. Stenson and T. M. Hickey, *Understanding Irish Spelling: A Handbook for Teachers and Learners – Tuiscint ar Litriú na Gaeilge: Lámhleabhar do Mhúinteoirí agus d'Fhoghlaimoirí*. Baile Átha Cliath: An Chomhairle um Oideachas Gaeltachta agus Gaelscolaíochta, 2019.
- [14] A. Bürki, P. Welby, M. Clément, and E. Spinelli. “Orthography and second language word learning: Moving beyond ‘friend or foe?’”. *Journal of the Acoustical Society of America*, vol. 145, pp. EL265–271, 2019.
- [15] P. Welby, E. Spinelli, and A. Bürki. “Spelling provides a precise (but sometimes misplaced) phonological target. Orthography and acoustic variability in second language word learning,” *Journal of Phonetics*, vol. 94, 101172, 2022.
- [16] J. Steele, “Assessing the role of orthographic versus uniquely auditory input in acquiring new L2 segments,” Paper presented at the 7èmes Journées internationales du réseau français de phonologie, Aix-en-Provence, France, 2005.
- [17] E. Gouraya, F. Gouraya, F. Kombouare, and J. Vernaudeau (dir.), *Propositions d'écriture du nâa kwényi (langue parlée à l'extrême sud de la Grande Terre de la Nouvelle Calédonie et à l'Île des Pins)*. Nouméa: Académie des Langues Kanak, 2011.
- [18] Académie des Langues Kanak, *Propositions d'écriture du paici*. Nouméa: Académie des Langues Kanak, 2013.
- [19] Académie des Langues Kanak, *Propositions d'écriture du nâa numèè (Langue parlée dans la région de Yaté, Nouvelle-Calédonie). Kô yè tici ngê nâa numèè*. Nouméa: Académie des Langues Kanak, 2015.
- [20] Académie des Langues Kanak, *Propositions d'écriture du méré a' a'jië. Ké yu méré a' a'jië (langue parlée à Houailou, Bourail, Poya, Kouaoua et Ponérihouen)*. Nouméa: Académie des Langues Kanak, 2018.
- [21] Académie des Langues Kanak, *Propositions d'écriture du naa drubea (langue parlée dans les régions de Païta, et Unia à Yaté, Nouvelle-Calédonie). Kô yere tici ngi naa drubea*. Nouméa: Académie des Langues Kanak, 2018.
- [22] J.-C. Rivierre, *Dictionnaire cémuhî-français, suivi d'un lexique français-cémuhî*. Paris: Peeters, 1994.
- [23] B. Bigi. “A phonetization approach for the forced-alignment task in SPPAS,” *Human Language Technology. Challenges for Computer Science and Linguistics. Lecture Notes in Computer Science*, vol. 9561, pp. 515–526, 2016.
- [24] A. Nedjar-Guerre, G. Wattelez, C. Serra-Mallol, S. Frayon, and O. Galy. “Adolescent screen time and unhealthy food consumption in the context of the digital development in New Caledonia,” *PLoS One*, 2023. doi: 10.1371/journal.pone.0285374.
- [25] S. Geneix-Rabault. “Les actions de l'Académie des Langues Kanak en faveur de la production d'outils et de supports en et sur les langues kanak,” In *L'école plurilingue en outre-mer : Apprendre plusieurs langues, plusieurs langues pour apprendre : l'école plurilingue*, I. Nocus, J. Vernaudeau, and M. Paia, (eds); Presses Universitaires de Rennes, 2014.
- [26] V. Fillol and C. Colombel-Teuira. “Comparons nos répertoires et nos langues : premiers pas d'une éducation au plurilinguisme,” <https://www.forumlecture.ch>, 2021.
- [27] S. Atti, C. Bouanou, H. Diahioue, J.-P. Diahioue, M. Grynagier, P. Hnacipan, Y. Lepigeon, and P. Prévaux; M. Waheo (illustr.), “Et toi, comment-dis tu?” Nouméa: Centre Territorial de Recherche et de Documentation Pédagogique, 1995.
- [28] M. Candelier. Toutes les langues à l'école ! L'éveil aux langues, une approche pour la Guyane ? In *Pratiques et représentations linguistiques en Guyane : regards croisés*, pp. 369–386, I. Léglise and B. Migge (eds), Paris: IRD, 2007.
- [29] V. Fillol and J. Vernaudeau. Toutes les langues à l'école : Enseignement des langues kanak et éveil aux langues de la région Asie-Pacifique à l'école calédonienne. In F. Laroussi and F. Liénard (eds), *Plurilinguisme, politique linguistique et éducation : Quels éclairages pour Mayotte ?* Presses universitaires de Rouen et du Havre, 2011.

### Online resources

#### Accord de Nouméa ‘Nouméa Agreement’

Accord sur la Nouvelle-Calédonie signé à Nouméa le 5 mai 1998  
<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT00000055817>

#### Examples of online phonetizers:

##### Occitan:

<https://fonetizaire.locongres.com/>

##### English:

<https://glotdojo.com/phonetic>

##### French:

<https://easypronunciation.com/en/french-phonetic-transcription-converter>