

Pinchah Kristang: A Dictionary of Kristang

Luis Morgado da Costa

Interdisciplinary Graduate School – Global Asia,
Nanyang Technological University, Singapore
lmorgado.dacosta@gmail.com

Abstract

This paper describes the development and current state of *Pinchah Kristang* – an online dictionary for Kristang. Kristang is a critically endangered language of the Portuguese-Eurasian communities residing mainly in Malacca and Singapore. *Pinchah Kristang* has been a central tool to the revitalization efforts of Kristang in Singapore, and collates information from multiple sources, including existing dictionaries and wordlists, ongoing language documentation work, and new words that emerge regularly from relexification efforts by the community. This online dictionary is powered by the Princeton Wordnet and the Open Kristang Wordnet – a choice that brings both advantages and disadvantages. This paper will introduce the current version of this dictionary, motivate some of its design choices, and discuss possible future directions.

Keywords: kristang, online dictionary, wordnet, portuguese-malay creole, endangered language

1. Introduction

Pinchah Kristang (lit. to cast a net over Kristang) was officially launched, in Singapore, at the First Kristang Language Festival, in May 2017, by then Singapore’s Deputy Prime Minister, Mr. Teo Chee Hean. This Festival celebrated the culmination of over a year of successful language revitalization efforts by *Kodrah Kristang* (lit. Awaken, Kristang) – a grassroots initiative with the goal of revitalizing Kristang in Singapore.

Pinchah Kristang is an open-source online bilingual dictionary (English-Kristang) and it is powered by the Open Kristang Wordnet – an ongoing project, supported by an eclectic team of trained linguistics, heritage speakers, students and community volunteers.

The main goals of this dictionary include both documenting and helping disseminate this beautiful endangered language. Building this dictionary is a great way to better understand, document and revitalize Kristang and, by making it accessible through an online interface, it has also been a great way to give back and empower the community to remember and use their language.

The remainder of this paper will start by providing an introduction to Kristang and its revitalization efforts in Singapore, followed by a discussion of *Pinchah Kristang*’s stages of development. It will conclude with some notes on the current state of the project, followed by a discussion of concerns and future directions.

1.1. Kristang

Kristang is a critically endangered creole language, spoken mainly by Portuguese-Eurasian communities in Malacca and Singapore. Although estimates concerning the number of speakers is a sensible topic, it is generally agreed that the number of speakers does not go beyond the lower thousands, in both Singapore and Malaysia.

According to Pillai et al. (2017), in Malacca, Kristang is estimated to be spoken fluently by only about half of the residents at the Portuguese Settlement, which has approximately 800–1000 residents. In Singapore, where the situation is much direr, it is estimated that fewer than 100 fami-

lies use Kristang daily, and only a very small percentage is still passing it down to younger generations. It is also generally agreed that the number of speakers in both countries has seen a steady decline (Nunes, 1999; Martens Wong, 2017). Kristang is known by many different names, including: Bahasa Geragau, Bahasa Serani, Luso-Malay, Malacca Creole, Malaccan, Malaqueiro, Malaquenho, Malaquense, Malaquês, Malayo-Portuguese, Malaysian Creole Portuguese, Papia Cristao, Papia Kristang, Portuguese Patois, Português de Malaca, and Serani.¹ For the remainder of this paper, we will refer to it as Kristang. Kristang, referring to both the language and its people (Hancock, 2009; Baxter, 2005), is the name most recognized by the community speaking this language in Singapore, where this project had its birth.

Kristang is originally derived from Malay and Portuguese, having its roots in the beginning of the 16th century with the arrival of the Portuguese to Malacca. At the same time, most likely caused by the inflow of people and cultural exchange through the Portuguese maritime exploration, there is evidence to suggest that Kristang is also related to number of other Portuguese Creoles, sharing commonalities with other languages and creoles from Africa, India, Southeast Asia, and Southern China (Nunes, 1999; Fernández, 2012; Baxter, 2012; Nunes, 2012)

Since the fall of Malacca to the Dutch, in the mid-17th century, the Kristang community has survived through roughly three more centuries of colonial occupation by both the Dutch and the English empires. This was accompanied by a mixture of other languages that also influenced Kristang. According to a few studies (Baxter and Bastos, 2012; Pillai et al., 2015), traces of Chinese, Indian, Malay, Dutch, Sri Lankan, Filipino and English language elements are evident in Kristang.

Kristang is a Subject-Verb-Object (SVO) language, with its vocabulary largely derived from Portuguese, with heavy influence from Malay and light influence from several other languages, notably Dutch and English. Its grammar and phonology are closely related to Malay.

¹<https://www.ethnologue.com/language/mcm>



Figure 1: Screenshot of *Pinchah Kristang*'s Welcome Page

1.2. *Kodrah Kristang*: Revitalization in Singapore

Pinchah Kristang was developed within the larger context of *Kodrah Kristang*, an initiative to revitalize Kristang in Singapore. *Kodrah Kristang* (lit. Awaken, Kristang) is an initiative founded by Kevin Martens Wong, that started to take form in early 2015. The sole purpose of this initiative is to work with the community to reawaken the language and bring it back to a healthy level of use.

Kodrah Kristang has published and maintains a revitalisation plan² that describes the long-term plans, goals and projects to revitalize the Kristang language in Singapore, up to the year 2045 CE. Following this plan, the *Kodrah Kristang* team has launched a number of initiatives including free classes for children and adults, freely available audio and vocabulary courses, musical performances in Kristang, public talks, social media presence, as well as TV and radio interviews. These initiatives have reached thousands of people around the world, and have also greatly increased public awareness of this endangered language in Singapore.

1.2.1. *Jardinggu*: Relexification Effort

A specific project that is closely related to this dictionary is the relexification effort led by *Kodrah Kristang*, known as *Jardinggu* – from the concatenation of *jarding* (garden) and *linggu* (language), meaning ‘Language Garden’.

The primary goal of *Kodrah Kristang* is to encourage more people to learn and speak Kristang. Many young people, however, do not see the value in learning a language that is not ready for the modern age, and that lacks words for concepts that are important today, like ‘website’, ‘cell phone’ and ‘wifi’. Similar to what has happened in other languages, such as Hawaiian or Alutiiq (Kimura and Hawaiian Lexicon Committee, 2009), *Kodrah Kristang* created a language incubator that engages and invites the community to actively create new words that are missing from Kristang’s known lexicon. New word suggestions are presented to the com-

munity, who votes to accept or reject them as part of the language. This is a flexible and continuous process. And, ultimately, only a long term analysis will be able to determine which words were picked up by speakers.

However, in an effort to both document and facilitate the use of newly coined words, the Open Kristang Wordnet (and, by default, *Pinchah Kristang*) includes these words with a temporary status. These words have a special status within the dictionary and may, at a later stage, be removed or substituted by different words that receive greater acceptance.

2. Development of *Pinchah Kristang*

Pinchah Kristang's online interface is a simple web application powered by Python, Flask,³ and SQLite3. It is heavily inspired by the Open Multilingual Wordnet (Bond and Foster, 2013, OMW), and its interface is being built in tandem with the most recent OMW codebase⁴ – reusing many of its core structure and functionalities, including the database design, but completely diverging in its user interface.

Figure 1 shows the welcome page of *Pinchah Kristang*. The online dictionary uses a simple bilingual search interface, where both English and Kristang can be used in searches. In addition, it is also enabled by SQLite3's GLOB clause, allowing simple regular expression searches such as the use of wildcards (e.g. the search ‘cat*’ will match any words that start with ‘cat’).

2.1. The Dictionary Data

At its core, *Pinchah Kristang* is powered by the Princeton Wordnet (Fellbaum, 1998, PWN), along with two other wordnets - the Open Kristang Wordnet and the Extended Kristang Wordnet. The reason for using two Kristang wordnets is simple: creating and curating a high quality wordnet is extremely time consuming. The Open Kristang Wordnet is a standard wordnet, being developed by linguists and trained volunteers. On the other hand, the Extended Kristang Wordnet is a lemma-to-lemma dictionary mapping dis-

²<https://kodrahkristang.files.wordpress.com/2017/02/kaminyudi-kodramintu-v2.pdf>

³<https://flask.palletsprojects.com/en/1.1.x/>

⁴<https://github.com/globalwordnet/OMW/>

guised in the form of a wordnet to be able to share the same online platform. In the following sections, the origin and structure of the dictionary data will be discussed in further detail.

2.2. Data Gathering

Given the endangered and fragmentary status of Kristang, there is a lack of competent speakers from whom to draw consistent lexicographic knowledge. As such, our dictionary is actively collating and validating this necessary lexicographic knowledge using multiple different sources. These sources include:

1. paper dictionaries and word-lists made available through both published works (including but not limited to: Baxter and de Silva (2004), Marbeck (2011) and Scully and Zuzarte (2004)); and through families or individuals who have kept notes about the language;
2. linguistic publications concerning Kristang, which often contain either wordlists or elicited language data that has been glossed (see, for example, Hancock (1970), Baxter (1988), and Avram (2013));
3. language documentation work, lead by *Kodrah Kristang* or its affiliates. This includes work done by Michael Erlewine and his EL3212 2017 course in Field Methods in Linguistics (which focused on Kristang), at the National University of Singapore. This course piloted the data collection and entry into this wordnet using OMWedit (Morgado da Costa and Bond, 2015);
4. new words and senses that are produced by *Jardingu*, the Kristang lexical incubation project, introduced above, in Section 1.2.1;

2.3. The Open Kristang Wordnet

The Open Kristang Wordnet was built using the ‘expansion approach’ (Vossen, 1998), where the structure of another wordnet is used as pivot. In this approach, building a wordnet is essentially a translation effort – conserving the structure of the pivot wordnet and translating individual nodes of the hierarchy, which can easily be done incrementally (i.e. usually starting by a subset of frequent concepts). The Princeton Wordnet (Fellbaum, 1998) is, by far, the most frequently used pivot for wordnet projects around the world. As previously mentioned, this project drew large inspiration on the work of OMW – possibly the best example of the benefits provided by developing wordnets using the ‘expansion approach’. For many years, the OMW has linked dozens of open wordnets using PWN as the pivot structure. However, a change to the way the OMW operates was recently introduced with the creation of the Collaborative Interlingual Index (Bond et al., 2016, CILI) – an open, language agnostic, flat-structured index that links wordnets across languages without imposing the hierarchy of any single wordnet. Through CILI, multiple projects are now able to link to each other and to contribute directly to the set of CILI’s concepts without the penalty of being frozen within an imposed structure.

Perhaps not surprisingly, CILI was initially created using the full set of concepts provided by the PWN (i.e. all PWN

concepts have a direct link to CILI). As such, the quickest and easiest way to link a new wordnet to CILI is still to use the expansion approach with PWN’s hierarchy as pivot – and this is what the Open Kristang Wordnet is doing.

2.3.1. Linking and Validation

With the help of all data sources described above, a master word list of aligned Kristang and English lemmas was compiled (and continues to be updated regularly). Depending on the source, extra information such as part-of-speech, full definitions and alignments to other languages such as Portuguese and Malay is, sometimes, available. Currently, this master list contains 15,435 word alignments between English and Kristang. It is important to note, however, that many of these alignments are not unique, as different sources often provide the same or similar information.

The first step to create our wordnet was done using a custom script to project all available English alignments onto PWN and generate a list of all candidate senses based on the data we have collected. This script is able to leverage multiple levels of information (e.g. language alignments, part-of-speech, number of overlaps per concept, etc.) to generate a ranked list of candidate senses. A full description of this script is, unfortunately, outside the scope of this paper.

A total of 51,077 candidate senses, spanning over 19,445 unique synsets were generated using the process described above. This data has, since then, been in a continuous process of validation by multiple members of *Kodrah Kristang* core team and some of its higher fluency students. This validation process consists of determining, with a high degree of certainty, if a candidate sense is attested in Kristang or not. Each candidate sense has three possible states: rejected (i.e. there is no evidence that this sense is available in the language); accepted (i.e. this sense has been confirmed to exist in the language, albeit with varying degrees of usage frequency); undetermined (i.e. it is possible that this sense exists in the language, but there is not enough evidence to confirm this at the moment, and requires further documentation work).

Out of the 51,077 candidate senses generated by our method: 8,382 (~16.4%) have been rejected; 7,011 (~13.7%) have been accepted; and 3,692 (~7.2%) were left undetermined – a total of 19,085 (~37.4%) candidate senses have been checked.

Table 1 provides a summary of the data currently contained in the Open Kristang Wordnet. It currently includes slightly more than 5,300 synsets, with just over 7,000 senses. About 97.5% of all concepts have been hand-linked to CILI (i.e. 2.5% do not have a mappable concept in CILI).

POS	Synsets	%	Words	%	Senses	%
Noun	2,969	55.3	2,248	61.5	3,860	55.1
Verb	1,195	22.3	426	11.6	1,447	20.6
Adjective	1,005	18.7	834	22.8	1,435	20.5
Adverb	171	3.2	127	3.5	231	3.3
Non-ref	26	0.5	22	0.6	38	0.5
	5,366		3,657		7,011	

Table 1: Statistics for Open Kristang Wordnet

The Open Kristang Wordnet is currently supported both

in the new WN-LMF format⁵ and the tab-separated-value (TSV) format used by the original OMW specifications.

2.4. The Extended Kristang Wordnet

As mentioned above, building and curating a wordnet can be extremely time consuming. The Open Kristang Wordnet is only roughly one third of the way from checking the 51,000 generated candidate senses. And even though the order with which senses are checked tries to maximise the revitalization efforts (e.g. giving priority to checking senses that are currently used in the Kristang classrooms), using only one third of the data would greatly diminish the usefulness of the online dictionary.

The Extended Kristang Wordnet solves this issue by creating a flat wordnet linking all English-Kristang lemma pairs that were not yet covered by the Open Kristang Wordnet manual efforts. Strictly speaking, the Extended Kristang Wordnet is only a wordnet in the sense that it makes use of the WN-LMF format to create English-Kristang word pairings. This wordnet is not linked to CILI (or any other wordnet) and, as such, it has no hierarchy, and is unable to leverage on PWN's data to further populate the dictionary with data other than basic English lemmas (e.g. full English definitions are not available).

Currently, the Extended Kristang Wordnet introduces 6,887 extra concepts, linking 6,883 Kristang Lemmas to 7,972 English Lemmas. Since these word-pairs come from various sources of data, most often from simple wordlists, the Extended Kristang Wordnet does not contain parts-of-speech information – making it very difficult to provide any kind of structured statistics for this resource.

As the manual efforts of the Open Kristang Wordnet continue to manually check all the data, the size of the Extended Kristang Wordnet will slowly decrease. The purpose of the Extended Kristang Wordnet is merely to bridge coverage issues, and it will cease to exist as soon as the Open Kristang Wordnet finishes the validation process of the existing data.

3. Current state of *Pinchah Kristang*

This dictionary project is still in active development, and it is primarily maintained by one of the core members of *Kodrah Kristang*. It is important to note that both *Kodrah Kristang* and this dictionary are fully run on a voluntary basis and, as such, it is most definitely slower than what would be expected from funded/staffed projects.

Currently, *Pinchah Kristang* is a strictly bilingual dictionary (see discussion below about plans to further expand this). Whenever available, English data is provided by the PWN. Despite being powered by wordnets, *Pinchah Kristang* is currently hiding some of the more complex features available in a wordnet (e.g. the rich semantic hierarchy) – which is done with user friendliness in mind. Most users are either learners or heritage speakers trying to remember forgotten words. As such, overloading the dictionary with linguistic information irrelevant to these users would impose a toll on usability.

Nevertheless, the fact that the dictionary is powered by wordnets is still somewhat clear. Let's compare Figure 2 and Figure 3 below:

Figure 2 shows the search results for the lemma 'gatu' ('cat', in Kristang). Since this lemma is completely unambiguous, only one entry is shown. In this case, this entry has been linked to the PWN. The Kristang lemma is shown to the left (highlighted in yellow), and it is accompanied by the English lemmas provided by the PWN. The part-of-speech and the definition are also provided.

Figure 3, on the other hand, shows the search results for the lemma 'cat'. These results are a bit more verbose. The PWN presents nine senses for the lemma 'cat' – including the senses for the domestic cat, and the verbal sense synonymous with 'to vomit'. Both these entries have a linked Kristang sense, which appear on the left column ('gatu' and 'gumitah', respectively). The seven other English sense of the lemma 'cat' do not have any Kristang lemmas linked to them. This is currently shown by a question mark in place of the Kristang lemma. Clicking on this question mark on will redirect the user to *KlaiFalah* (lit. 'how to say'), an initiative under *Jardinggu*, the Kristang Lexical Incubator, where users can request and suggest new words to be created. In this platform, users are able to inquire about senses that are still missing from the dictionary. These senses might already exist in the language, but the data was not yet collected, or it might be considered by the lexical incubator program – which works with the community to create a new word for the missing sense. As this process is done outside the scope of this dictionary project, it will not be discussed here in further detail.

Lastly, Figure 3 shows a tenth sense (in the second line) with is a match for the lemma 'cat' but in Kristang. One of the existing data sources lists *cat* as a Kristang word for 'pain'. However, as it can be seen from the lack of part-of-speech and definition, this Kristang lemma belongs to the Extended Kristang Wordnet. As of this moment, this use is not yet attested beyond a written source and, as such, it has not been linked to the PWN. This is also a good example to show how the single search input form matches both English and Kristang lemmas.

4. Concerns and Future Directions

This section will focus on specific points of concern for this project, and outline some of the future directions the project is likely to take.

4.1. Non-English Concepts

A problem that is common to many wordnets that followed the 'expand' approach, is the difficulty in breaking away from the preexisting structure of the pivot project – in our case, CILI / the Princeton Wordnet.

It is not surprising, therefore, that we share these same concerns. There are many concepts that are not currently represented in Open Kristang Wordnet for the simple reason that these concepts do not have an English counterpart in the PWN. Some of these concepts are culturally specific, such as *kari debal* or *pang susi* – two of the most charismatic recipes of Kristang cuisine – and would not be expected to exist in PWN. But some other basic concepts are also missing due to way Kristang lexifies certain concepts that are treated compositionally in English. This happens, for example, with *dosora* (two o'clock) and *desora* (ten o'clock) –

⁵<https://github.com/globalwordnet/schemas>

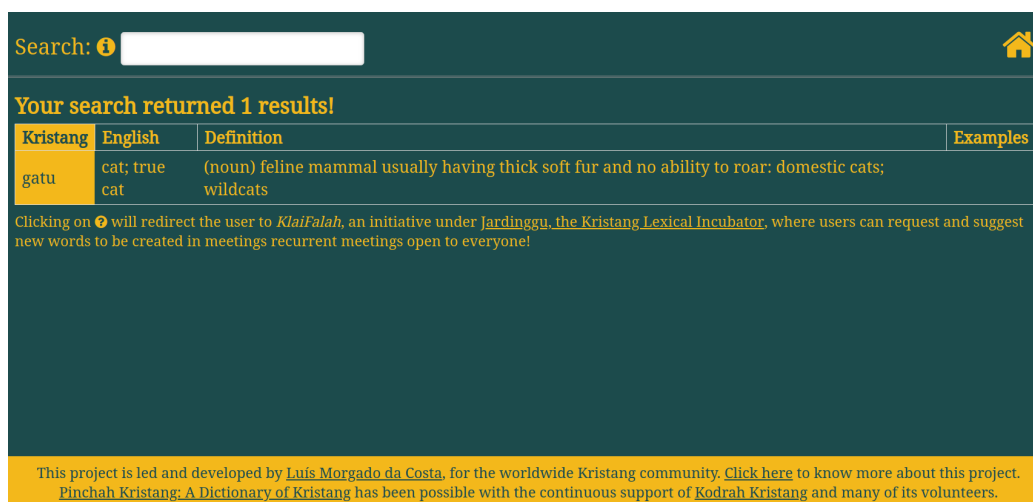


Figure 2: Screenshot of Pinchah Kristang’s Search Results for ‘gatu’



Figure 3: Screenshot of Pinchah Kristang’s Search Results for ‘cat’

which are deemed compositional in English (ten + o’clock), and hence do not have a concept in PWN.

Currently, this problem is mitigated by the use of the Extended Kristang Wordnet – which provides these Kristang words and translations to the dictionary, but lacks more structured data such as the part-of-speech or a definition.

Moving forward, the solution for this is for this project to be fully integrated within CILI (Bond et al., 2016), and gain the ability to suggest new concepts that necessary to adequately represent Kristang.

4.2. Relexification

Another topic of concern to this dictionary project is the decision to support the relexification efforts of *Kodrah Kris-*

tang, which has been briefly introduced above, in Section 1.2.1. Even though supporting these efforts without the guarantee that a word has permanently entered the language may seem rash at first glance, it is important to note that *Pinchah Kristang*’s main goal is to support Kristang’s revitalization. In addition, for all practical purposes, these new words enter the language once they are voted in by the community – even if with a temporary status. And despite the real chance that some of these words might end up substituted by new words in the future, these new words may still see usage before they are deprecated, in favor of different words. One of *Pinchah Kristang*’s goals is also to help disseminate these new words, with the hope of updating Kristang’s vocabulary to modern times.

Ultimately, *Jardinggu* is a parallel effort towards Kristang’s revitalization, with its own procedures and coordination. And lists of newly created words, notes on how the words were created, and their respective voting results are stored outside this dictionary. For this reason, and even though these new words hold a special status in the dictionary data, the current interface does not show the provenance of any of its senses (including words created by the relexification efforts). This decision might, however, be revised in the future – organically accompanying and responding to the ways the lexical incubator operates.

4.3. Examples and Corpus Integration

Pinchah Kristang has also started working on further supporting the documentation and revitalization of Kristang through the creation of a sense-tagged example corpus. Even though a wordnet is capable, in principle, to better disambiguate different senses and their usages (i.e. when compared to a simple bilingual word-list, which is the form of most dictionaries), this is not always enough. Having an example corpus, that provides users with clear usage examples of each sense is, to some extent, essential to full-fledged dictionary projects. Many wordnet projects have pioneered this line of thinking, and built their wordnets in parallel with an example corpus or, in the best case scenario, a sense-tagged example corpus. *Pinchah Kristang* will follow in their footsteps, and is currently working on a sense-tagged example corpus to enrich its user-experience.

4.4. Orthographic Variation

Since Kristang has rarely been recorded in written form over the course of its history, it never acquired a standard written form broadly accepted by its multiple communities. Orthographic and phonetic influences from geographical or cultural proximity with Kristang communities has resulted in wide orthographic variation in the language. Portuguese, Malay and English orthographies have all left a mark on Kristang orthography in Singapore. For example, Baxter and de Silva (2004) lists ‘dog’ as *kachoru*, where the word-medial consonant /tʃ/ follows the Portuguese or English orthographic system, while Marbeck (2011) lists it as *kacoru*, with the same /tʃ/ consonant spelled as it would be in Malay – both phonetically equivalent. Similarly, *kuelu* (Marbeck, 2011) and *kwelu* (Scully and Zuzarte, 2004) for ‘rabbit’ closely resembles Malay orthography, while the variant spelling *coelho* (Scully and Zuzarte, 2004) undoubtedly mimics Portuguese orthography.

The prevalence of metathesis in certain consonant clusters is another related problem. Examples of this include ‘-dr-’ (*kodrah* and *kordah* for ‘to wake up’), ‘-tr-’ (*kotri* and *korti* for ‘town’ or ‘city’), and ‘-br-’ (*ebra* and *erba* for ‘grass’). These are not, however, phonetically equivalent – posing a new level of complexity that should, at least, be acknowledged by *Pinchah Kristang*.

Pinchah Kristang aims to recognize all variant spellings included in existing dictionaries as well as others widely used by speakers in Singapore. Currently, these are all listed as different senses. In the future *Pinchah Kristang* will move towards the representation of a canonical form with multiple spelling variations. The canonical form will be orthography

used by *Kodrah Kristang*, which follows the orthography proposed by Baxter and de Silva (2004). Only senses that have the same phonetic realization will be merged (e.g. *kachoru* and *kacoru*, meaning ‘dog’) but not in cases of metathesis (e.g. *ebra* and *erba*, both meaning ‘grass’). The current reasoning to deal with metathesis in a different way than plain orthographic variation comes from the fact that metathesis introduces phonetic differences – which is more closely tied to the identity of the speakers. And while we believe there is little hurt in moving towards standardization of the written form (since most of the variation comes from the fact that a written form never really existed), standardizing phonetic variation would impose stronger claims over how the language should be used. We want to move away from politicizing the language as much as possible.

4.5. Pronunciation and Phonetic Representation

In keeping with the role of supporting the documentation and revitalization efforts of Kristang, *Pinchah Kristang* will soon also include both voice recordings and phonetic representation (i.e. IPA) for some or all of our wordnet senses. These efforts are being pursued in tandem with language documentation interviews, that have provided hundreds of individual, segmented sense pronunciation recordings (sometimes by more than one speaker).

4.6. Linked Etymology

Largely inspired by projects such as the Etymological Wordnet (Melo, 2014) and the World Loanword Database (Haspelmath and Tadmor, 2009, WOLD), it would be extremely interesting to leverage on CILI and other linked wordnets (through the OMW) to study and integrate information concerning possible etymologies of Kristang lemmas in *Pinchah Kristang*.

Currently, the OMW includes large wordnets for the two most important languages from which Kristang vocabulary derives, Portuguese (de Paiva and Rademaker, 2012) and Malay (Mohamed Noor et al., 2011), along with a number of wordnets for other languages that are known to have contributed to the Kristang lexicon, namely English (Fellbaum, 1998), Dutch (Postma et al., 2016), Mandarin Chinese (Wang and Bond, 2013) and Cantonese (Sio and Costa, 2019). Missing from this list are, unfortunately, languages such as Hokkien, Hindi, Tamil and Sanskrit – all known to have influenced Kristang’s lexicon.

The plan to use other wordnets to explore and link etymologies of Kristang senses could end up making our bilingual dictionary into a small multilingual dictionary, where languages of interest could be shown alongside relevant senses. Since this would most certainly be extremely time consuming, it would perhaps be interesting to employ semi-automatic methods to measure similarity between Kristang senses and all languages of interest.

4.7. Towards a Multilingual Dictionary

In line with what was discussed in the section above, even prior to having full-fledged etymological links between wordnets, the possibility of adding both Portuguese and Malay as parallel data in the dictionary is currently under discussion. Strictly from a dictionary standpoint, this

would be relevant, as it would allow other communities (namely Portuguese and Malay speaking communities) to use the dictionary and get acquainted with Kristang. However, from a revitalization point-of-view, this might also bring certain challenges concerning language identity and ‘purity’. As discussed in Section 4.4, on orthographic variation, certain dictionaries (and this is also true for individual people within the community) like to align themselves to either Malay or Portuguese spelling (even when this introduces orthographic ambiguity or a more opaque orthography). Language and identity cannot easily be dissociated, and missteps concerning these topics might impose constraints towards healthy revitalization of Kristang. *Pinchah Kristang* is currently monitoring this possibility, and it will align its design to best serve the Kristang community.

5. License and Release Notes

This dictionary is freely available online⁶ and its main objective is to help the documentation and dissemination efforts of *Kodrah Kristang*, by collating information that would otherwise be dispersed and at a risk of being lost.

We want to encourage others to use our work, to work towards the further improvement of Kristang and its communities, to avoid replication of efforts, and to inspire other endangered communities to work towards their language preservation. For this reason, all work developed for *Pinchah Kristang: A Dictionary of Kristang and the Open Kristang Wordnet* are developed under a Creative Commons Attribution 4.0 International⁷ (CC BY 4.0) license.

6. Acknowledgments

Ultimately, this is a project for and by the Kristang community (and their friends). And the current state of the dictionary and the wordnet would not have been possible without the volunteered hours of many who dedicated their time to this project. In particular, we would like to thank the many volunteers who helped with many forms of data entry and validation⁸, as well as all permanent members and participants of the *Jardinggu* meetings; we would also like to thank Michael Erlewine and his 2017 class of EL3212 Field Methods in Linguistics, at the National University of Singapore, for piloting data collection and entry into this wordnet.

7. Bibliographical References

- Avram, A. A. (2013). The Dutch lexical contribution to three Asian Portuguese Creoles // Contribuição lexical do holandês para três crioulos de base portuguesa da Ásia. *PAPIA-Revista Brasileira de Estudos do Contato Linguístico*, 23(1):51–74.
- Baxter, A. N. and Bastos, A. (2012). A closer look at the post-nominal genitive in asian creole portuguese. *Ibero-Asian creoles: Comparative perspectives*, 46:47–79.

- Baxter, A. N. and de Silva, P. (2004). *A dictionary of Kristang (Malacca Creole Portuguese) with an English-Kristang finderlist*. Number 564, Series B. Pacific Linguistics, The Australian National University, Australia.
- Baxter, A. N. (1988). *A Grammar of Kristang (Malacca Creole Portuguese)*. Number 95, Series B. Pacific Linguistics, The Australian National University, Australia.
- Baxter, A. N. (2005). Kristang (Malacca Creole Portuguese) –A long-time survivor seriously endangered. *Estudios de sociolingüística*, 6(1):1–37.
- Baxter, A. N. (2012). Vestiges of etymological gender in malacca creole portuguese. *Pidgins and Creoles in Asia*, 38:115–149.
- Bond, F. and Foster, R. (2013). Linking and extending an open multilingual wordnet. pages 1352–1362, Sofia.
- Bond, F., Vossen, P., McCrae, J. P., and Fellbaum, C. (2016). Cili: the collaborative interlingual index. In *Proceedings of the Global WordNet Conference*, volume 2016.
- de Paiva, V. and Rademaker, A. (2012). Revisiting a Brazilian wordnet. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*, Matsue.
- Christiane Fellbaum, editor. (1998). MIT Press, Cambridge, MA.
- Fernández, M. A. (2012). Nenang, nino, nem não, ni no. *Ibero-Asian Creoles: Comparative Perspectives*, 46:205–237.
- Hancock, I. F. (1970). Some dutch-derived items in Papia Kristang. *Bijdragen tot de Taal-, Land-en Volkenkunde* 126, (3):352–356.
- Hancock, I. F. (2009). The portuguese creoles of malacca. *Revue roumaine de linguistique*, 54(3-4):295–306.
- Martin Haspelmath et al., editors. (2009). *WOLD*. Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Kimura, L. and Hawaiian Lexicon Committee. (2009). Indigenous new words creation perspectives from alaska and hawai ‘i. *Indigenous Language Revitalization: Encouragement, Guidance and Lessons Learned*, pages 121–139.
- Marbeck, J. M. (2011). *Commemorative Bahasa Serani Dictionary*. Adil Enterprise, Malaysia.
- Martens Wong, K. (2017). Bos Papiah Kristang? (Do You Speak Kristang?): A Eurasian Linguistic Legacy. In *Singapore Eurasians: Memories, Hopes and Dreams*, pages 369–379. World Scientific.
- Melo, G. D. (2014). Etymological wordnet: Tracing the history of words. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, Reykjavik, Iceland, May. European Language Resources Association (ELRA).
- Mohamed Noor, N., Sapuan, S., and Bond, F. (2011). Creating the open Wordnet Bahasa. In *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation (PACLIC 25)*, pages 258–267, Singapore.
- Morgado da Costa, L. and Bond, F. (2015). Omwedit - the integrated open multilingual wordnet editing system. In *Proceedings of the 53rd Annual Meeting of the Associ-*

⁶https://www.luismc.com/pinchah_kristang/

⁷<https://creativecommons.org/licenses/by/4.0/>

⁸A full list of contributors to the wordnet and online dictionary can be found here: https://www.luismc.com/pinchah_kristang/about

- ation for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, System Demonstrations (ACL 2015), pages 73–78, Beijing, China.
- Nunes, M. (1999). The use of kristang in the portuguese settlement of malacca. *Journal of Modern Languages*, 12:147–158.
- Nunes, M. P. (2012). Traces of superstrate verb inflection in makista and other asian-portuguese creoles. *Ibero-Asian Creoles: Comparative Perspectives*, 46:289–326.
- Pillai, S., En, C. M., and Baxter, A. N. (2015). Vowels in malacca portuguese creole. *Research in Language*, 13(3):248–265.
- Pillai, S., Phillip, A., and Soh, W.-Y., (2017). *Revitalizing Malacca Portuguese Creole*, pages 1–17. Springer International Publishing, Cham.
- Postma, M., van Miltenburg, E., Segers, R., Schoen, A., and Vossen, P. (2016). Open Dutch WordNet. In *Proceedings of the Eight Global Wordnet Conference*, Bucharest, Romania.
- Scully, V. and Zuzarte, C. (2004). *The Most Comprehensive Eurasian Heritage Dictionary: Kristang-English / English-Kristang*. SNP International, Singapore.
- Sio, J. U.-S. and Costa, L. M. D. (2019). Building the Cantonese Wordnet. In *Proceedings of the Tenth Global Wordnet Conference*, Wroclaw, Poland.
- Vossen, P. (1998). A multilingual database with lexical semantic networks. *Dordrecht: Kluwer Academic Publishers. doi*, 10:978–94.
- Wang, S. and Bond, F. (2013). Building the chinese open wordnet (cow): Starting from core synsets. In *Sixth International Joint Conference on Natural Language Processing*, pages 10–18.