

## Contact languages around the world and their levels of endangerment

Nala H. Lee

*National University of Singapore*

This paper provides an up-to-date report on the vitality or endangerment status of contact languages around the world, including pidgins, creoles, and mixed languages. By utilizing information featured in the *Endangered Languages Project* and the *Atlas of Pidgin and Creole Languages* online portals, 96 contact languages are assessed on the Language Endangerment Index, a method of assessment that is based on four factors including intergenerational transmission, absolute number of speakers, speaker number trends, and domains of use. Results show that the contact languages are most at risk with respect to intergenerational transmission and domains of use. This is explained by the social and historical nature of contact languages. Overall results further raise the concern that the proportion of pidgins, creoles and mixed languages at some level of risk is extremely high. Reasons are provided for why linguists should be concerned about the endangerment of these languages.

**1. Introduction**<sup>1</sup> While the language endangerment problem has been generally well-highlighted in linguistics, the endangerment of contact languages has not received the same level of attention. Krauss (1992) postulates that at least half and possibly as much as 90% of the world's languages will no longer be spoken by the end of the present century, while a more recent empirical study estimates a slightly less catastrophic rate of loss, at one language every three months (Campbell et al. 2013). At any rate, the threat of language extinction is undeniable, with consequences ranging from threats to linguistic diversity and the potential loss of part of the sum of human knowledge (Hale 1992), to threats to the well-being and identity of speakers. In response to these threats, the subfields of language documentation and conservation have made rapid advancements. However, most of the language endangerment discussion neglects the subset of languages that comprises pidgins, creoles, mixed languages, and other forms of contact languages. Apart from individual documentation and revitalization projects that target specific contact languages, such as the documentation of Sri Lanka Malay (Ansaldo & Lim 2014) and Michif revitalization efforts (Iseke 2013), as well as the survey of endangered Pacific pidgins and creoles (Erhart & Mühlhäusler 2007), there has been very little said about the endangerment and loss of contact languages as a collective whole. In effect, contact languages are

<sup>1</sup>This paper has benefitted from the valuable comments and suggestions provided by the two anonymous reviewers, and by the editor of *Language Documentation & Conservation*, Nicholas Thieberger. Any errors that remain are solely mine.

often neglected in research on language endangerment, based on the rather dubious and narrowly defined criterion that these languages do not exhibit sufficiently unique linguistic features.<sup>2</sup> Rather, there has been a “pervasive tendency to privilege strongly autonomous languages over those that self-evidently (i.e. lexically) derive from existing languages” (Garrett 2006:182). This has had repercussions in terms of funding or the lack thereof from international agencies dedicated to the documentation and preservation of endangered languages (see Garrett 2006). There is hence an impetus to highlight the threat that these contact languages face in reality, and the value of preserving them.

The notion that much more can be said about how pidgins, creoles, mixed languages, and other forms of contact languages feature in the language endangerment discussion is not new. Garrett (2006:175) states that much of the language endangerment discussion focuses on “relatively obscure and ‘exotic’ languages, and almost never contact languages”, reiterating Mühlhäusler’s (2003:243–244) stance in a review of books concerning language endangerment, that “none of the authors addresses the massive loss of Pidgins, Creoles and other impure contact languages ... We need to beware of the narrowly focused discursive construction of endangered languages”. With specific regard to language documentation efforts, O’Shannessy (2012) states that not much attention has been paid to new contact varieties, due to the perceived need to document older, disappearing traditional languages. The same notion also underlies Mufwene’s (2017) recent statement that the discourse in the subfield of language endangerment and loss has left out nonindigenous populations, including “European settlement colonies”, “Bantu settled territories”, and “plantation settlement colonies” (see also Lee 2017). This lack of inclusion is problematic, particularly so when pidgins, creoles, and other forms of contact languages are said to be doubly endangered or “doubly marginalized” (Garrett 2012:145) for various reasons, including the fact that contact languages are often perceived as “merely ‘broken’ or ‘corrupted’ versions of the European languages to which they are self-evidently related” (Garrett 2012:148). As an example, a common sentiment among Hawai’i Creole speakers is that their language, also called Pidgin, is not a real language. In the same vein, Valdman (1987:107) highlights the precariousness of creoles specifically, stating that “[c]reoles are inexorably destined to dissolve in these major languages via the process of decreolization” (see also DeGraff 2001). It then becomes important to establish the exact magnitude of the language endangerment problem for contact languages. One way of doing so is to provide an assessment of the levels of endangerment (or vitality) of various contact languages around the world. The goal of this paper is to provide an up-to-date report on the endangerment statuses of contact languages.

It has become possible to provide a vitality assessment of languages using a newer method of assessment that generates a score for the level of endangerment regardless of much is known of a language (unless there is absolutely no appropriate information available). The *Language Endangerment Index* (LEI) was developed for use with the *Catalogue of Endangered Languages* (ELCat), hosted on the *Endangered Languages*

<sup>2</sup>This comment must be credited to one of the anonymous reviewers.

*Project* platform ([www.endangeredlanguages.com](http://www.endangeredlanguages.com)). The LEI aims to provide a level of endangerment for any language, based on four criteria: intergenerational transmission, absolute number of speakers, speaker number trends, and domains of use. By using a multi-factored method of evaluation that is not based solely on speaker numbers, LEI addresses an issue faced by those who have had their research projects on contact languages rejected by funders on the basis that speaker numbers were not sufficiently reduced. LEI also differs from its predecessors, such as the EGIDS (Lewis & Simons 2010) and UNESCO's nine factors for the assessment of language vitality (UNESCO 2003), in its ability to provide a level of certainty alongside an overall score indicating the language's level of endangerment (Lee & Van Way 2016). The level of certainty is based on the number of factors that are utilised in the assessment. The more factors utilised in the assessment, the more certain one can be about the level of endangerment of the language. Crucially, such a mechanism ensures that a vitality score can be assigned, even if there is only one factor for which there is information. This addresses a major concern that there is simply not enough known about many of the world's languages. Lehmann (1999) suggests that it is probable that nothing is known about half of the world's languages, aside from their names. While this is most likely an overstatement, the type of information required for assessing language vitality is usually more limited or not available at all. LEI then becomes appropriate for use when the aim of the evaluation is to provide a comparison of the vitality of a large number of languages, as in the case of ELCat. LEI is therefore suitable for the objective of providing vitality assessments for a comprehensive list of contact languages.

The mechanisms and motivations behind the design of LEI are explained in Lee & Van Way (2016). For this paper, the following information suffices. Languages are assessed on four scales, including intergenerational transmission (whether the language is being passed on to younger generations), absolute number of speakers, speaker number trends (whether numbers are increasing, decreasing, or stable), and domains of use (whether the language is used in a wide number of domains or limited ones). These scales are attached to this paper in the Appendix. Languages are scored on each scale, and each individual score correlates with a descriptor. If a language scores 0 on a scale, it is regarded as *safe* on the scale. If it scores a 5 on the scale, it is regarded as being *critically endangered* on the scale. The descriptors for scores 1, 2, 3, and 4 are *vulnerable*, *threatened*, *endangered*, and *severely endangered* in ascending order of risk. Individual scores for each factor are then aggregated, with the score for intergenerational transmission multiplied by two, since intergenerational transmission is the most essential factor for the continued survival of any language. The total score for all factors used is converted to a percentage of the highest attainable score. If all four factors are used, the highest attainable score is 25, since each factor has a maximum score of 5, except for intergenerational transmission which is doubly weighted and has a maximum possible score of 10. If only two factors are used (excluding intergenerational transmission), the highest attainable score is 10. The exact method for deriving overall level of endangerment is represented by the following formula:

Level of endangerment =  $\{[(\text{intergenerational transmission score} \times 2) + \text{absolute number of speaker score} + \text{speaker number trends score} + \text{domains of use score}] / \text{total possible score based on number of factors used}\} \times 100$

Results generated are then interpreted using the index on the left in Table 1. Scores are paired with discrete overall levels of endangerment. For example, if a language scores between 81% and 100%, it is assessed to be *critically endangered*. The scale on the right provides a certainty score based on the number of factors used in the assessment. If a language is assessed solely based on intergenerational transmission, the total number of points possible would be 10, and if the language is assessed based on two factors (excluding intergenerational transmission), the total number of possible points would still be 10, which indicates that the assessor can be 40% certain about the vitality assessment. Both endangerment and certainty scores should be considered together for a more accurate overview of any language's degree of endangerment.

**Table 1.** Language Endangerment Index and levels of certainty

Language Endangerment Index	Level of certainty
100-81% = Critically endangered	25 points = 100% certain, based on the evidence available for each of the four factors
80-61% = Severely endangered	20 points = 80% certain, based on the evidence available
60-41% = Endangered	15 points = 60% certain, based on the evidence available
40-21% = Threatened	10 points = 40% certain, based on the evidence available
20-1% = Vulnerable	5 points = 20% certain, based on the evidence available
0% = Safe	

The languages selected for assessment are languages for which up-to-date sociological information is more readily available. These are languages featured on ELCat and on the *Atlas of Pidgins and Creole Languages Online* (APiCS) ([apics-online.info](http://apics-online.info)). ELCat aims to be an online platform for up-to-date resources on endangered languages, and it reflects information on the four factors required for the LEI assessment where obtainable. Based on the available information, ELCat then assigns the language a LEI vitality score. At the time of writing, ELCat provides vitality information on 31 contact languages: Barikanchi, Berbice Creole Dutch, Bungi, Chinook Wawa, Gibanawa, Iha Based Pidgin, Javindo, Kodiak Russian Creole, Louisiana Creole, Malabar Sri-Lanka Portuguese, Malaccan Creole Portuguese, Nauru Pacific Pidgin, Ndyuka-Trio Pidgin, Ngatik Men's Creole, Nicaragua Creole English, Kinubi, Palenquero, Macao Creole Portuguese, Petjo, Norf'k, San Miguel Creole French, Set-tla, Sri Lanka Malay, Torres Strait Creole, Yilan Creole, Copper Island Aleut, Kallaway, Mbugu, Michif, Tagdal, and Wutunhua. To provide a fuller understanding of the state of contact languages around the world, languages featured on APiCS are also assessed for their individual levels of endangerment. In addition to provid-

ing typological information on the contact languages included, APiCS also features sociological information that allows for vitality assessments. Where information is available regarding intergenerational transmission, speaker numbers, speaker number trends, and domains of use, the language is assessed on the relevant individual scales (appended to this paper). For languages featured in both resources, the resource with the more complete and updated information is utilized. For example, the information on APiCS for Norf'k is utilized since APiCS provides information about speaker numbers, speaker number trends, and domains of use, whereas ELCat only has information about Norf'k's absolute number of speakers. The total number of contact languages surveyed is thus brought up to 96.

Note that the vitality or endangerment of these 96 contact languages are discussed as a coherent whole. While some attempt could be made to subcategorize these languages into the various types of contact languages (such as pidgins, creoles, or mixed languages), there are a number of languages that may not typically fit into these categories (for example, see Markey 1982 on whether Afrikaans is a creole; Mufwene 2009 on why it is not necessary to treat Kikongo-Kituba as a creole; Ansaldo & Lim 2014 on why it more appropriate to classify Sri Lanka Malay as a restructured variety of Malay rather than as a creole). It is also not the aim of this article to compare between different varieties of contact languages with the goal of indicating which among these is most at risk. Instead, the article aims to shed light on the fact that contact languages, as a significant category on their own, warrant attention in wider discussions on language endangerment.

The following sections provide information on how the contact languages fare on the individual scales of intergenerational transmission, absolute number of speakers, speaker number trends, and domains of use. An overall survey of language endangerment of these languages is then provided,<sup>3</sup> followed by a discussion on what these numbers portend for the future of contact languages, and why linguists should care.

**2. Contact languages on the scale of intergenerational transmission** Intergenerational transmission is doubly-weighted on the LEI, for the reason that without transmission to the next generation, a language will cease to exist naturally, regardless of how many people who are not children speak it, or regardless of how many domains it is used in (see Lee & Van Way 2016).

Table 2 shows how the contact languages surveyed fare on the scale of intergenerational transmission. Only languages for which there was available and appropriate information on intergenerational transmission on ELCat or APiCS are represented. The higher the score given on the scale, the more likely the language is not being transmitted to younger speakers, and a score of zero would entail that the language is *safe* in this regard – that the language is spoken by all members of the community including children. The individual scales for the interpretation of these results are attached in the Appendix.

<sup>3</sup>Preliminary overall results are mentioned in Lee (2017) in a wider discussion on the endangerment on nonindigenous languages. The results presented here have been updated, and are complete and far more extensive.

**Table 2.** Contact languages on the scale of intergenerational transmission; 0-safe, 1-vulnerable, 2-threatened, 3-endangered, 4-severely endangered, 5-critically endangered.

Language	Scale of Intergenerational Transmission
Baba Malay	5
Batavia Creole	dormant
Berbice Creole Dutch	dormant
Bungi	dormant
Cape Verdean Creole of São Vicente	0
Cavite Chabacano	4
Chinese Pidgin English	dormant
Chinese Pidgin Russian	dormant
Copper Island Aleut	dormant
Eskimo Pidgin	dormant
Ghanaian Pidgin English	2
Javindo	5
Louisiana Creole	4
Macao Portuguese Creole	5
Malaccan Creole Portuguese	3
Mauritian Creole	0
Ndyuka-Trio Pidgin	dormant
Negerhollands	dormant
Ngatik Men's Creole	4
Nicaragua Creole English	0
Pidgin Hawaiian	dormant
Reunion Creole	0
Santome	1
Singapore Bazaar Malay	3
Singlish	0
Sranan	2
Sri Lanka Malay	1
Torres Strait Creole	1
Yilan Creole	2
Yimas-Arafundi Pidgin	5

Out of the 96 contact languages surveyed, 10 are entirely *dormant*. The term *dormant* is used for any language that is thought to have lost its last remaining speaker recently (within approximately the last fifty years), or for any language where any doubt remains as to the possible existence of any speakers (Lee & Van Way 2016). That a language is *dormant* does imply that it is no longer passed down through any form of intergenerational transmission. *Dormant* languages not passed on include Ndyuka-Trio Pidgin, Chinese Pidgin English, Chinese Pidgin Russian, Pidgin Hawaiian, Eskimo Pidgin, Bungi, Negerhollands, Batavia Creole, Berbice Creole Dutch, and

Copper Island Aleut. There are 20 languages for which there is information regarding intergenerational transmission, and for which such a scale is appropriate.

Out of these 20 languages, 4 are *critically endangered* on the scale of intergenerational transmission: Baba Malay, Javindo, Macao Portuguese Creole, and Yimas-Arafundi Pidgin. These languages are only spoken among a limited subset of people from the grandparents' generation or older. Another three languages are *severely endangered* on the same scale: Cavite Chabacano, Louisiana Creole, and Ngatik's Men's Creole. Another two languages are *endangered*<sup>4</sup> on the scale of intergenerational transmission: Malaccan Creole Portuguese and Singapore Bazaar Malay. On the same scale, three languages are *threatened* (Ghanaian Pidgin English, Sranan, and Yilan Creole) and another three are *vulnerable* (Santome, Sri Lanka Malay, and Torres Strait Creole). Only five languages are considered *safe* on this list: Cape Verdean Creole of São Vicente, Mauritian Creole, Nicaragua Creole English, Reunion Creole, and Singlish. For the remaining 66 languages, there is no information pertaining to intergenerational transmission in the sources utilized. For example, while there are other types of social information available on Babalia Creole Arabic (Häberl in press) and Korlai (Clements 2013), such as speaker numbers, there is no precise information concerning intergenerational transmission for these languages in ELCat or APiCS. Languages that are not assessed on the scale of intergenerational transmission also include languages for which such a scale is not appropriate. For example, Gurindji Kriol is a mixed language spoken by mostly Gurindji people under the age of 40, while older Gurindji people speak Gurindji, often codeswitching with Kriol (Meakins 2013). Given that the older generation speaks Gurindji and Kriol, but not Gurindji Kriol, the regular notion of intergenerational transmission would not apply in the case of Gurindji Kriol. The scale of intergenerational transmission, that assumes that elderly speakers would be the last bastions of a language, would not be appropriate for use with this language at the current point in time.

As the results demonstrate, languages that are *safe* on the scale of intergenerational transmission account for only 5.2% of the total number of languages surveyed (5 out of 96 languages). Intergenerational transmission remains a key factor in the continued survival of any language, and contact languages are particularly susceptible to losing intergenerational transmission. Many of these contact languages developed alongside a more dominant language, having emerged in the context of colonial expansion, displacement, and relocation, and still exist alongside a more dominant language (Garrett 2006; Lee 2017). For those working on an endangered language, regardless of whether it is a contact or non-contact language, these are familiar circumstances. Colonialism itself and imbalances of power often factor into language endangerment (see Grenoble & Whaley 1998 for multiple examples). For a contact language, the dominant language is more often than not one of the source languages in the case of a contact language. With a cline resulting between the dominant source language and the contact variety, and what Garrett (2012:148) terms as a "lack of autonomy", speakers may think that they are speaking the same language as the prestigious variety (Mufwene 2003). This inevitably has repercussions on intergen-

<sup>4</sup>*Endangered* is the LEI label denoting mid-level language endangerment.

erational transmission, when the perceived target is the dominant source language. Speakers may also be pressured not to pass on the contact language, viewing the inherent choice of transmission to be the dominant source language, which is often regarded as being socioeconomically beneficial. A contact language that is spoken alongside its lexifier language is Singlish. Singlish is used in Singapore, where the official language of administration and education is English, and parents are officially discouraged from speaking Singlish to their children as part of a Speak Good English Movement (Velayutham 2007). While the campaign seems to have no effect on intergenerational transmission for now, and the language is still *safe* on the scale of intergenerational transmission, there is no foretelling what may happen in the future. Any adverse effect on intergenerational transmission would be dire for this language. Contact languages are naturally also susceptible to the pressures of dominant languages that are not their source languages (Garrett 2006; see also Lee 2017). For example, once spoken on the islands of St. Thomas, St. John, and St. Croix (the current US Virgin Islands), Negerhollands was eventually passed over in favor of English. Younger speakers began acquiring English, and the Dutch-based creole lost its last speaker in 1987 (van Sluijs 2013).

In effect, intergenerational transmission has been recognized to be so vital to the continuity of any language that it is specifically targeted by revitalization programs. Language nests, a concept that first originated in New Zealand for the revitalization of Māori, refers to early childhood education (in pre-school, day care, or crèche contexts) entirely grounded in the target language and accompanying culture (Benton 2015). While the concept is no longer new, it is one that has only recently been taken up for contact languages, for example in the form of Kakakihkih Kristang, a play-group in Singapore for children aged between 4 and 6 that is conducted entirely in Kristang (Wong, p.c.), also known as Malaccan Creole Portuguese.

**3. Contact languages on the scale of absolute number of speakers** Another factor considered on the LEI is the absolute number of speakers that a language has. A criticism of reliance on this criterion is that speaker numbers may not be the best indicator of a language's viability. For example, languages such as the major Quechua languages, with millions of speakers, can be endangered, with speakers viewing the language as an impediment to social mobility (Adelaar 2014). As has been repeatedly noted in the literature, a language with over a million speakers, none of whom are children, is endangered, since the language will not survive into the future (see discussion in §2). However, speaker number is oftentimes the only information about vitality available for a language, and it is evident that it is much more difficult to envision a viable future for languages with fewer than 10 speakers as opposed to a language with more than 100,000 speakers.

Table 3 shows how the languages fare on the LEI's scale of absolute number of speakers. Only languages for which there was speaker number information on ELCat or APiCS are represented.

**Table 3.** Contact languages on the scale of absolute number of speakers; 0-*safe*, 1-*vulnerable*, 2-*threatened*, 3-*endangered*, 4-*severely endangered*, 5-*critically endangered*.

Language	Speaker numbers	Scale of Absolute Number of Speakers
Afrikaans	>6,000,000	0
Ambon Malay	200,000	0
Angolar	5,000	2
Baba Malay	2000	2
Babalia Creole Arabic	3,940	2
Bahamian Creole	250,000	0
Batavia Creole	dormant	dormant
Belizean Creole	150,000	0
Berbice Creole Dutch	dormant	dormant
Bislama	<1,000,000	0
Bungi	dormant	dormant
Cameroon Pidgin English	>1,000,000	0
Cape Verdean Creole of Brava	6,000	2
Cape Verdean Creole of Santiago	450,000	0
Cape Verdean Creole of São Vicente	76,000	1
Casamancese Creole	10,000	1
Cavite Chabacano	4,000	2
Chinese Pidgin English	dormant	dormant
Chinese Pidgin Russian	dormant	dormant
Chinook Wawa	<10	5
Copper Island Aleut	dormant	dormant
Diu Indo-Portuguese	180	3
Eskimo Pidgin	dormant	dormant
Fad'Ambô	5,000	2
Fanakalo	3,000,000	0
Ghanaian Pidgin English	5,000,000	0
Guadeloupean Creole	600,000	0
Guinea-Bissau Kriyo	600	3
Gullah	<10000	2
Gurindji Kriol	1,000	2
Guyanais	64,000	1
Guyanese Creole	700,000	0
Haitian Creole	11,000,000	0
Hawai'i Creole	~700000	0
Iha Based Pidgin	5,500	2
Jamaican	3,000,000	0
Javindo	<100	4
Juba Arabic	<1,000,000	0
Kallawaya	<99	4
Kikongo-Kituba	10,000,000	0
Korlai	780	3
Kriol	>20,000	1
Lingala	15,000,000	0
Louisiana Creole	<7000	2
Macao Portuguese Creole	~50	4
Malabar-Sri Lanka Portuguese	4	5
Malaccan Creole Portuguese	~2150	2

Continued from previous page

Language	Speaker numbers	Scale of Absolute Number of Speakers
Martinican Creole	600,000	0
Mauritian Creole	>1300000	0
Mbugu	7,000	2
Media Lengua	<10,000	2
Michif	<100	4
Nauru Pacific Pidgin	<10,000	2
Ndyuka-Trio Pidgin	dormant	dormant
Negerhollands	dormant	dormant
Nengee	66,500	1
Ngatik Men's Creole	15-30	4
Nicaragua Creole English	35,000-50,000	1
Nigerian Pidgin	~75,000,000	0
Norf'k	800	3
Palenquero	2000	2
Papiamentu	300,000	0
Petjo	<100	4
Pichi	>100,000	0
Pidgin Hawaiian	dormant	dormant
Principense	<100	4
Reunion Creole	800,000	0
San Andres Creole English	20,000	1
San Miguel Creole French	3	5
Sango	2,500,000	0
Santome	60,000	1
Saramaccan	50,000	1
Seychelles Creole	100,000	0
Sierra Leone Krio	350,000	0
Singapore Bazaar Malay	<10,000	2
Singlish	>4,000,000	0
Sranan	126,000	0
Tagdal	<20,000	1
Tayo	3,000	2
Ternate Chabacano	3,000	2
Tok Pisin	>3,000,000	0
Torres Strait Creole	~6000	2
Trinidad English Creole	1,200,000	0
Vincentian Creole	117,000	0
Wutunhua	4,000	2
Yilan Creole	~3000	2
Yimas-Arafundi Pidgin	<5	5
Zamboanga Chabacano	300,000	0

There are only 7 languages out of the 96 surveyed for which there is no speaker number information available on either ELCat or APiCS. These are Barikanchi, Gibanawa, Kinubi, Kodiak Russian Creole, Pidgin Hindustani, Settla, and Sri Lanka Malay. Aside from the 10 languages that are *dormant* and have no known speakers, 4 are

*critically endangered* with fewer than 10 speakers each at the last count. These are Chinook Wawa, San Miguel Creole French, Malabar-Sri Lanka Portuguese, and Yimas Arafundi Pidgin. Another 7 languages are *severely endangered* with fewer than 100 speakers each: Javindo, Kallawaya, Macao Portuguese Creole, Michif, Ngatik Men's Creole, Petjo, and Principense. With less than 1,000 speakers each, another 4 languages are *endangered* on the scale of absolute number of speaker: Diu Indo-Portuguese, Guinea-Bissau Kriyo, Korlai, and Norf'k. A total of 31 languages are *threatened* or *vulnerable* on the same scale. *Threatened* languages have between 1,000 and 9,999 speakers, while *vulnerable* languages have between 10,000 and 99,999 speakers. Languages that can be considered to be *safe*, at least on the scale of absolute number of speakers, account for only 34.4% of the total number of languages surveyed (33 out of 96 languages). These languages have at least 100,000 speakers.

Again, while a language's vitality is not completely dependent on speaker numbers, speaker numbers are important when they are the only information available on a language. Among the 96 languages surveyed, there are 8 for which no other type of information apart from absolute number of speakers is available for the assessment of language vitality. These are Babalia Creole Arabic, Chinook Wawa, Iha Based Pidgin, Kallawaya, Nauru Pacific Pidgin, San Miguel Creole French, Malabar-Sri Lanka Portuguese, and Tagdal. The assessments of these languages can thus be viewed with 20% certainty, since one factor (that is not intergenerational transmission) is utilized. Hence, languages such as Chinook Wawa, San Miguel Creole French, and Malabar-Sri Lanka Portuguese are *critically endangered* at a 20% level of certainty. For other languages for which there is more information, it becomes necessary to utilize these other forms of information as well.

**4. Contact languages on the scale of speaker number trends** Absolute number of speakers is a static reflection of language vitality. For a more dynamic approach to speaker numbers, it is crucial to consider speaker number trends. The criterion of speaker number trends in LEI addresses whether speaker numbers for a given language are increasing, decreasing, or relatively stable. At the *critically endangered* end of the scale, a small percentage of the community speaks the language, and speaker numbers are decreasing very rapidly. At the *safe* end of the same scale, almost all community members speak the language, and speaker numbers are stable or increasing.

Table 4 demonstrates how the 96 surveyed languages fare on the scale of speaker number trends.

**Table 4.** Contact languages on the scale of speaker number trends; 0-*safe*, 1-*vulnerable*, 2-*threatened*, 3-*endangered*, 4-*severely endangered*, 5-*critically endangered*.

Language	Scale of Speaker Number Trends
African American English	1
Afrikaans	1
Ambon Malay	0

*Continued from previous page*

Language	Scale of Speaker Number Trends
Angolar	1
Baba Malay	5
Bahamian Creole	0
Batavia Creole	dormant
Belizean Creole	0
Berbice Creole Dutch	dormant
Bislama	0
Bungi	dormant
Cameroon Pidgin English	1
Cape Verdean Creole of Brava	0
Cape Verdean Creole of Santiago	0
Cape Verdean Creole of São Vicente	0
Casamancese Creole	2
Cavite Chabacano	5
Chinese Pidgin English	dormant
Chinese Pidgin Russian	dormant
Copper Island Aleut	dormant
Diu Indo-Portuguese	1
Eskimo Pidgin	dormant
Fad'Ambô	0
Fanakalo	1
Ghanaian Pidgin English	0
Guadeloupean Creole	0
Guinea-Bissau Kriyo	0
Gullah	1
Gurindji Kriol	0
Guyanais	0
Guyanese Creole	0
Haitian Creole	0
Hawai'i Creole	1
Jamaican	0
Javindo	5
Juba Arabic	1
Kikongo-Kituba	0
Kinubi	3
Kodiak Russian Creole	5
Korlai	0
Kriol	1
Lingala	0
Louisiana Creole	4
Macao Portuguese Creole	5
Malaccan Creole Portuguese	4
Martinican Creole	0
Mauritian Creole	0
Mbugu	1
Media Lengua	2
Michif	5
Ndyuka-Trio Pidgin	dormant
Negerhollands	dormant

*Continued from previous page*

Language	Scale of Speaker Number Trends
Nengee	0
Ngatik Men's Creole	4
Nicaragua Creole English	0
Nigerian Pidgin	0
Norf'k	1
Palenquero	0
Papiamentu	0
Petjo	5
Pichi	0
Pidgin Hawaiian	dormant
Pidgin Hindustani	2
Principense	5
Reunion Creole	0
San Andres Creole English	2
Sango	0
Santome	4
Saramaccan	1
Seychelles Creole	0
Sierra Leone Krio	0
Singapore Bazaar Malay	5
Singlish	0
Sranan	1
Sri Lanka Malay	4
Tayo	0
Ternate Chabacano	4
Tok Pisin	0
Torres Strait Creole	1
Trinidad English Creole	1
Vincentian Creole	0
Yimas-Arafundi Pidgin	5
Zamboanga Chabacano	0

Among the 96 contact languages surveyed, there are 13 languages for which there is no information on speaker number trends. These are Babalia Creole Arabic, Barikanchi, Chinook Wawa, Gibanawa, Iha Based Pidgin, Kallawaya, Nauru Pacific Pidgin, San Miguel Creole French, Settla, Malabar-Sri Lanka Portuguese, Tagdal, Wutunhua, and Yilan Creole. Speaker number trends are also not applicable to the 10 *dormant* languages. An additional ten languages are *critically endangered* on the scale of speaker number trends: Baba Malay, Cavite Chabacano, Javindo, Kodiak Russian Creole, Macao Portuguese Creole, Michif, Petjo, Principense, Singapore Bazaar Malay, and Yimas-Arafundi Pidgin. These languages are spoken by a small percentage of the community, and speaker numbers are decreasing rapidly. Another 6 languages are *severely endangered* on the same scale: Louisiana Creole, Malaccan Creole Portuguese, Ngatik Men's Creole, Santome, Sri Lanka Malay, and Ternate Chabacano. One language is assessed to be *endangered*: Kinubi. A total of 4 languages are as-

essed to be *threatened*: Casamancese Creole, Media Lengua, Pidgin Hindustani, and San Andres Creole English, and 16 more are assessed as *vulnerable*: African American English, Afrikaans, Angolar, Cameroon Pidgin English, Diu Indo-Portuguese, Fanakalo, Gullah, Hawai'i Creole, Juda Arabic, Kriol, Mbugu, Norf'k, Saramaccan, Sranan, Torres Strait Creole, and Trinidad English Creole. The remaining 36 languages are *safe* on the scale of speaker number trends, with almost all of the community speaking the language and stable or increasing speaker numbers.

In all, the proportion of contact languages surveyed that is *safe* on the criterion of speaker number trends is 37.5% (36 out of 96 languages). Conversely, the proportion of contact languages that is at some level of risk on this scale (after subtracting *safe* and *dormant* languages, and languages for which no information is available) is 38.5%, with 37 out of 96 languages being rated from *vulnerable* to *critically endangered*.

**5. Pidgin, creoles and mixed languages on the scale of domains of use** The last criterion on the LEI is that of domains of use. It operates on the notion of Fishman's *domains of language use*, which refers to "interactions that are rather unambiguously related (topically and situationally) to one or another of the major institutions of society" (Fishman 1991:44), and assumes that an implicit cline can be observed when language shift occurs (Lee & Van Way 2016). At a more advanced stage of language shift, the language is used in fewer domains and in informal ones. At a less advanced stage of language shift, the language is used in more domains, and in both formal as well as informal ones. Results relevant to domains of use are presented in Table 5.

**Table 5.** Contact languages on the scale of domains of use; 0-*safe*, 1-*vulnerable*, 2-*threatened*, 3-*endangered*, 4-*severely endangered*, 5-*critically endangered*.

Language	Scale of Domains of Use
African American English	2
Afrikaans	0
Ambon Malay	1
Angolar	3
Baba Malay	5
Bahamian Creole	1
Batavia Creole	dormant
Belizean Creole	2
Berbice Creole Dutch	dormant
Bislama	1
Bungi	dormant
Cameroon Pidgin English	2
Cape Verdean Creole of Brava	0
Cape Verdean Creole of Santiago	1
Cape Verdean Creole of São Vicente	1
Casamancese Creole	2
Cavite Chabacano	4
Chinese Pidgin English	dormant
Chinese Pidgin Russian	dormant

*Continued from previous page*

Language	Scale of Domains of Use
Copper Island Aleut	dormant
Diu Indo-Portuguese	3
Eskimo Pidgin	dormant
Fad'Ambô	1
Fanakalo	5
Ghanaian Pidgin English	2
Guadeloupean Creole	1
Guinea-Bissau Kriyo	0
Gullah	3
Gurindji Kriol	2
Guyanais	2
Guyanese Creole	1
Haitian Creole	1
Hawai'i Creole	1
Jamaican	1
Juba Arabic	1
Kikongo-Kituba	1
Kinubi	1
Korlai	3
Kriol	2
Lingala	1
Macao Portuguese Creole	5
Martinican Creole	2
Mauritian Creole	1
Media Lengua	2
Michif	4
Ndyuka-Trio Pidgin	dormant
Negerhollands	dormant
Nengee	1
Ngatik Men's Creole	4
Nicaragua Creole English	1
Nigerian Pidgin	1
Norf'k	2
Palenquero	2
Papiamentu	1
Pichi	3
Pidgin Hawaiian	dormant
Principense	4
Reunion Creole	1
San Andres Creole English	2
Sango	1
Santome	3
Saramaccan	2
Seychelles Creole	0
Sierra Leone Krio	1
Singapore Bazaar Malay	5
Singlish	2
Sranan	1
Sri Lanka Malay	4

Continued from previous page

Language	Scale of Domains of Use
Tayo	3
Ternate Chabacano	3
Tok Pisin	1
Trinidad English Creole	1
Vincentian Creole	1
Wutunhua	3
Yilan Creole	4
Yimas-Arafundi Pidgin	5
Zamboanga Chabacano	1

Of the 96 contact languages surveyed, there are 29 languages that cannot be assessed for domains of use. There are 19 languages for which there is no information of that type in ELCat or APiCS, as well as 10 *dormant* languages, for which these domains of use do not apply. Of the other 67 languages, 5 languages are *critically endangered*, meaning that they are used in very few specific domains, such as in ceremonies, songs, prayers and in limited domestic activities. These languages are Baba Malay, Fanakalo, Macao Portuguese Creole, Singapore Bazaar Malay, and Yimas-Arafundi Pidgin. Another 6 languages are *severely endangered*: Cavite Chabacano, Michif, Ngatik Men's Creole, Principense, Sri Lanka Malay, and Yilan Creole. An additional 9 languages are *endangered*: Angolar, Diu Indo-Portuguese, Gullah, Korlai, Pichi, Santome, Tayo, Ternate Chabacano, and Wutunhua. A total of 43 languages are *threatened* or *vulnerable*. Only 4 languages are assessed as *safe* on the scale of domains of use: Afrikaans, Cape Verdean Creole of Brava, Guinea-Bissau Kriyo, and Seychelles Creole. These languages are used in most domains, including official ones, such as government, mass media, and education. According to these numbers, only 4.2% of the contact languages surveyed (4 out of 96) are *safe* on the scale of domains of use. Conversely, after subtracting languages that are *safe* on the scale, *dormant* languages, and languages for which no domains of use information is available, 65.6% of the total number of languages surveyed (63 out of 96) is at some level of known risk.

It is unsurprising that contact languages generally do not fare well where domains of use are concerned. In fact, the numbers show that they are the least *safe* on the scale of domains of use (4.2%), as compared to all other factors: intergenerational transmission (marginally so at 5.2%), absolute number of speakers (34.4%), and speaker number trends (37.5%). The poor performance of these languages on the scale of intergenerational transmission was earlier explained by the fact that these languages usually exist alongside a more dominant language – there exists a choice of whether to pass on the contact language or the more dominant language to the younger generation, and this choice is often motivated by factors such as social mobility and economic opportunity. In relation to domains of use, contact languages are more often than not used in limited domains. For example, pidgins are said to have arisen as “socially ...restricted contact languages” (Thomason & Kaufman 1988:170). Velupillai (2015) identifies various types of pidgins, including trade and

nautical pidgins, workforce pidgins, military pidgins, and urban pidgins, all of which are associated with particular domains of use. Speakers of these languages usually have other languages that they speak in most domains, unless the contact language becomes the primary means of communication for the group. While other contact languages such as creoles are associated with wider functions, these languages have their own social and historical narratives that limit the potential of them having unfettered use in all domains. Given observations that “creole languages are popularly derided as corrupt and inferior forms of a “standard” language like English or French” (Frank 2007:1), creoles often function as an unofficial or low language in a diglossic type of relationship with an official language. For example, standard English is used for most official domains in places such as Jamaica and the Republic of Trinidad and Tobago, while Jamaican (or Jamaican Creole) and Trinidad English Creole function as unofficial vernaculars (Farquharson 2013; Mühleisen 2013). While these are languages that would typically not be perceived as being at risk, having more than a million speakers each (see Table 3), the fact of the matter is that a language that cannot be used in a full set of domains is never truly *safe*, and that it is easier for a language to lose domains than to gain them (Lee 2017). It is also notable that even creoles that are given some sort of national status, seldom have equal and unfettered use in all domains (Lee 2017). For example, Haitian Creole is seldom used in courts (DeGraff 2017) and Tok Pisin is used only in the first three years of formal education if the community has chosen it as a medium of education (Smith & Siegel 2013). Due to their intrinsic nature, contact languages such as creoles are used in curbed domains – it comes as no surprise that it is in this respect that they become most susceptible to language endangerment and loss.

**6. The overall levels of endangerment of contact languages** With individual scores given for intergenerational transmission, absolute number of speakers, speaker number trends, and domains of use, overall scores for level of endangerment and the level of certainty with which these assessments can be regarded are calculated for each language surveyed. Derived results for the 96 contact languages are presented in Table 6.

**Table 6.** Levels of endangerment and certainty of assessments of 96 contact languages

Language	Level of Endangerment	Level of Certainty (%)	Region	Source
African American English	vulnerable	60	North America	APICS
Afrikaans	vulnerable	60	Southern Africa	APICS
Ambon Malay	vulnerable	60	Southeast Asia	APICS
Angolar	threatened	60	West Africa	APICS
Baba Malay	critically endangered	100	Southeast Asia	ELCat
Babalia Creole Arabic	threatened	20	Central & East Africa	ELCat
Bahamian Creole	vulnerable	60	North America	APICS
Barikanchi	NA	NA	West Africa	ELCat
Batavia Creole	dormant	NA	Southeast Asia	APICS

*Continued from previous page*

Language	Level of Endangerment	Level of Certainty (%)	Region	Source
Belizean Creole	vulnerable	60	Caribbean	APICS
Berbice Dutch Creole	dormant	NA	Caribbean	ELCat
Bislama	vulnerable	60	Pacific	APICS
Bungi	dormant	NA	North America	ELCat
Cameroon Pidgin English	vulnerable	60	West Africa	APICS
Cape Verdean Creole of Brava	vulnerable	60	West Africa	APICS
Cape Verdean Creole of Santiago	vulnerable	60	West Africa	APICS
Cape Verdean Creole of São Vicente	vulnerable	100	West Africa	APICS
Casamancese Creole	threatened	60	West Africa	APICS
Cavite Chabacano	severely endangered	100	Southeast Asia	APICS
Chinese Pidgin English	dormant	NA	East Asia	APICS
Chinese Pidgin Russian	dormant	NA	East Asia	APICS
Chinook Wawa	critically endangered	20	North America	ELCat
Copper Island Aleut	dormant	NA	North America	ELCat
Diu Indo-Portuguese	endangered	60	South Asia	APICS
Eskimo Pidgin	dormant	NA	North America	APICS
Fad'Ambô	vulnerable	60	West Africa	APICS
Fanakalo	threatened	60	South Africa	APICS
Ghanaian Pidgin English	vulnerable	100	West Africa	APICS
Gibanawa	NA	NA	West Africa	ELCat
Guadeloupean Creole	vulnerable	60	Caribbean	APICS
Guinea-Bissau Kriyo	vulnerable	60	West Africa	APICS
Gullah	threatened	60	North America	APICS
Gurindji Kriol	threatened	60	Australia	APICS
Guyanais	vulnerable	60	Caribbean	APICS
Guyanese Creole	vulnerable	60	Caribbean	APICS
Haitian Creole	vulnerable	60	Caribbean	APICS
Hawai'i Creole	vulnerable	60	Pacific	APICS
Iha Based Pidgin	threatened	20	Southeast Asia	ELCat
Jamaican	vulnerable	60	Caribbean	APICS
Javindo	critically endangered	80	Southeast Asia	ELCat
Juba Arabic	vulnerable	60	Central & East Africa	APICS
Kallawaya	severely endangered	20	South America	ELCat
Kikongo-Kituba	vulnerable	60	Central & East Africa	APICS
Kinubi	threatened	40	Central & East Africa	APICS
Kodiak Russian Creole	critically endangered	20	North America	ELCat
Korlai	threatened	60	South Asia	APICS
Kriol	threatened	60	Pacific	APICS
Lingala	vulnerable	60	Central & East Africa	APICS
Louisiana Creole	severely endangered	80	North America	ELCat
Macao Portuguese Creole	critically endangered	100	East Asia	ELCat
Malabar-Sri Lanka Portuguese	critically endangered	20	South Asia	ELCat
Malaccan Creole Portuguese	endangered	80	Southeast Asia	ELCat
Martinican Creole	vulnerable	60	Caribbean	APICS

*Continued from previous page*

Language	Level of Endangerment	Level of Certainty (%)	Region	Source
Mauritian Creole	vulnerable	100	Indian Ocean	APICS
Mbugu	threatened	40	Central & East Africa	APICS
Media Lengua	threatened	40	South America	APICS
Michif	critically endangered	40	North America	APICS
Nauru Pacific Pidgin	threatened	20	Pacific	ELCat
Ndyuka-Trio Pidgin	dormant	NA	Caribbean	ELCat
Negerhollands	dormant	NA	Caribbean	APICS
Nengee	vulnerable	60	Caribbean	APICS
Ngatik Men's Creole	severely endangered	100	Pacific	ELCat
Nicaragua Creole English	vulnerable	100	Caribbean	APICS
Nigerian Pidgin	vulnerable	60	West Africa	APICS
Norfolk	threatened	60	Pacific	APICS
Palenquero	threatened	60	Caribbean	APICS
Papiamentu	vulnerable	60	Caribbean	APICS
Petjo	critically endangered	40	Southeast Asia	ELCat
Pichi	vulnerable	60	West Africa	APICS
Pidgin Hawaiian	dormant	NA	Pacific	APICS
Pidgin Hindustani	threatened	40	Pacific	APICS
Principense	critically endangered	60	West Africa	APICS
Reunion Creole	vulnerable	100	Indian Ocean	APICS
San Andres Creole English	threatened	60	Caribbean	APICS
San Miguel Creole French	critically endangered	20	Central America	ELCat
Sango	vulnerable	60	Central & East Africa	APICS
Santome	threatened	100		APICS
Saramaccan	threatened	60	Caribbean	APICS
Settla	NA	NA	Southern Africa	ELCat
Seychelles Creole	safe	60	Indian Ocean	APICS
Sierra Leone Krio	vulnerable	60	West Africa	APICS
Singapore Bazaar Malay	severely endangered	100	Southeast Asia	APICS
Singlish	vulnerable	100	Southeast Asia	APICS
Sranan	threatened	100	Caribbean	APICS
Sri Lanka Malay	endangered	80	Southeast Asia	ELCat
Tagdal	vulnerable	20	West Africa	ELCat
Tayo	threatened	60	Pacific	APICS
Ternate Chabacano	endangered	60	Southeast Asia	APICS
Tok Pisin	vulnerable	60	Pacific	APICS
Torres Strait Creole	vulnerable	80	Australia	ELCat
Trinidad English Creole	vulnerable	60	Caribbean	APICS
Vincentian Creole	vulnerable	60	Caribbean	APICS
Wutunhua	endangered	40	East Asia	ELCat
Yilan Creole	endangered	80	East Asia	ELCat
Yimas-Arafundi Pidgin	critically endangered	100	Pacific	APICS
Zamboanga Chabacano	vulnerable	60	Southeast Asia	APICS

Levels of endangerment are generated for 93 of the 96 surveyed contact languages. For the remaining three languages, no vitality information is available – nothing is known about whether there is intergenerational transmission, number of speakers, whether speaker numbers are increasing or decreasing, and what domains they are utilized in. These languages are Barikanchi, Gibanawa, and Settla. Among the 93 other languages, 10 are *dormant*, as mentioned above. Only one language, Seychelles Creole, is considered *safe*, but at a 60% level of certainty, based on information regarding absolute number of speakers, speaker number trends, and domains of use. This leaves 82 contact languages at some level of risk. A total of 11 languages are *critically endangered*: Baba Malay, Chinook Wawa, Javindo, Kodiak Russian Creole, Macao Portuguese Creole, Michif, Petjo, Principense, San Miguel Creole French, Malabar-Sri Lanka Portuguese, and Yimas-Arafundi Pidgin. An additional 5 languages are *severely endangered*: Cavite Chabacano, Kallawaya, Louisiana Creole, Ngatik Men's Creole and Singapore Bazaar Malay. Another 6 languages are *endangered*: Diu Indo-Portuguese, Malaccan Creole Portuguese, Sri Lanka Malay, Ternate Chabacano, Wutunhua, and Yilan Creole. A total of 21 languages are *threatened*: Angolar, Babalia Creole Arabic, Casamancese Creole, Fanakalo, Gullah, Gurindji Kriol, Iha Based Pidgin, Kinubi, Korlai, Kriol, Mbugu, Media Lengua, Nauru Pacific Pidgin, Norf'k, Palenquero, Pidgin Hindustani, San Andres Creole English, Santome, Saramaccan, Sranan, and Tayo. The remaining 39 languages are *vulnerable*.

What do these numbers portend for contact languages? How does the state of these numbers stack up against that of all languages around the world in general? Overall, the number of contact languages that are deemed to be *safe* form an extremely small fraction of the total number of languages surveyed, at 1.0% (1 out of 96 languages). A total of 10.4% (10 out of 96) are *dormant*. Another 11.5% (11 out of 96) are *critically endangered*, while 5.2% (5 out of 96) are assessed as *severely endangered*, and 6.3% (6 out of 96) are *endangered*. Another 21.9% (21 out of 96) are *threatened*, and 40.6% (39 out of 96) are *vulnerable*. Nothing is known about the vitality of the remaining 3.1% (3 out of 96) of the languages surveyed.

In all, the proportion of contact languages that is at some level of risk is high, at 85.4% (82 out of 96 languages), after the number of *safe* and *dormant* languages and the languages for which there is no vitality information are subtracted from the total number of languages surveyed. This falls within the high range of Krauss' (1992) prediction, that states that possibly as many as 90% of the world's languages will no longer be spoken by the end of the present century. The figure becomes even higher, if languages that are at risk of loss and languages that have already been lost in the last fifty years or so, are considered together. The proportion of contact languages that is at some level of risk or already *dormant* is 95.8% (92 out of 96 languages), after the number of *safe* languages and the languages for which no vitality information is available is subtracted from the total number of languages surveyed. At the time of writing, there are 3,407 languages that are at various levels of risk or that are already *dormant*. This is the total number of languages featured in the *Catalogue of Endangered Languages* ([www.endangeredlanguages.com](http://www.endangeredlanguages.com)). Given that there are about 6,879 languages around the world, after adjusting the 7,099 languages on

*Ethnologue* (Simons & Fennig 2017) for 219 languages with no known speakers and one constructed language, this means that 49.5% of the world's languages are either at risk or already *dormant*. What the comparison of numbers indicate then, is that the risk of endangerment and loss for contact languages is essentially almost twice as great as that for all the world's languages.

**7. Why care about the loss of contact languages?** The concern that contact languages could be “doubly marginalized” or doubly endangered was previously raised (Garrett 2012:145), and this current investigation substantiates that position. At twice the risk of endangerment and loss as languages around the world in general, the threat that contact languages face warrants urgent attention. The problem is further compounded, as noted above, by the fact that not enough attention is being paid to the endangerment and loss of these languages (see Mühlhäusler 2003; Garrett 2006; O’Shannessy 2012).

In addition, while often endangered by regionally dominant languages, contact languages are also born within the context of endangerment. Research on these languages must contend with the opinion that language shift may possibly be maladaptive, since it is language contact that leads to the emergence of contact languages, just as it is language contact that can lead to language death (see Mufwene 2017; Lee 2017). In reality, the relationship between language contact, language death, and the genesis of contact languages is less than straightforward. The perspective that contact languages necessarily represent a shift away from heritage languages towards dominant languages, such as English, is problematic. Construed from another point of view, contact languages represent a halfway point between the indigenous language and the lexifier language, a point at which language shift can be halted. At such a point, the contact language symbolizes a separate identity that is neither simply part of the lexifier group identity, nor a straightforward composite of the identities associated with the component languages. This is particularly evident in the case of mixed languages that emerge in contact situations wherein at least one of the early speaker groups is bilingual in the source languages (Thomason 1997). Mixed languages often emerge as acts of identity (see McConvell & Meakins 2005; O’Shannessy 2005). For example, the mixed language Light Warlpiri represents not only an establishment of a local identity within the wider Warlpiri speech community by the young Warlpiri, but also a resistance to shift away from Warlpiri itself (O’Shannessy 2005; 2012). A contact language, then, can also represent a way of partially maintaining a threatened language (see O’Shannessy 2012). The case of mixed languages therefore differs somewhat from that of pidgins and creoles, which may emerge in the absence of a lingua franca (see Thomason 1997). However, it is important to note that this does not imply that there is no identity function involved where these other forms of contact languages are concerned. Whereas the identity function emerges in tandem with the creation of mixed languages, creoles may develop this function later. In the development of plantation creoles for example, while enslaved or contract laborers shifted from their languages towards a dominant language, the creoles that ensued became a part of the speakers’ unique social identity. In the case of Hawai’i Creole,

the language has become emblematic of solidarity and local culture (Drager 2012). Any threat to such a contact language then, denotes a threat to cultural or ethnic identity. The loss of cultural or ethnic identity is recognized as one of the major consequences of language loss (Tsunoda 2005). Aside from this, there are also other compelling reasons for why more attention should be paid to the endangerment of contact languages.

All other conventionally discussed consequences of language loss apply to contact languages, just as they do to non-contact languages. Other consequences of language loss include the loss of part of the sum of human knowledge (Hale 1992), the loss of linguistic diversity (Hale 1992), and the loss of languages themselves, on which depends the ability of linguists to discover the full range of what is possible in human language and cognition (Lee & Van Way 2016). Some of these arguments have been presented in Lee (2017) but are reiterated here to emphasize what is at risk of being lost. Just as traditional languages encode information regarding local knowledge and ecosystems, such as how the Siberian Tofa have different and complex names for reindeer depending on their life stages (Harrison 2007), contact languages too can encode various types of knowledge about the world. Kallawaya, spoken in the highlands of Bolivia, is a *severely endangered* mixed language that encodes over 900 species of medicinal plants (Girault 1989). Aside from botanical and ecological knowledge, there are other types of cultural knowledge that is encoded in contact languages. In fact, it is postulated that the Baba Malay-speaking communities that were formed via early intermarriages between Chinese traders and indigenous women in the Malay Archipelago may have preserved ancestral worship better than in the Southern Chinese provinces, where the original male ancestry can be traced to, due to impact of the Cultural Revolution that affected China (Sankar et al. 2016). Now, with the language being *critically endangered*, cultural concepts such as the *datok dapur* ‘deity of the kitchen’ and ceremonies such as *pai tigong* ‘praying to the sky god’ are at risk of being lost. So, just as with the loss of any non-contact language, the world loses knowledge and ways of looking at itself when a contact language is lost.

The knowledge argument can and must be extended to linguistic knowledge. All studied languages make up the sum of what linguists know about language. Just as non-contact languages such as Yéli Dnye of Papua New Guinea provide evidence for double articulations (where a consonantal closure is made in more than one place), and languages such as Tzeltal of Mexico demonstrate that positionals are a valid word class (Evans & Levinson 2009), much can be obtained from contact languages. Linguists are particularly concerned about representing all language stocks in the world, including isolates, since only then will it be possible to understand the range of all that is possible in human languages (Whalen & Simons 2012). Arguably, contact languages that represent unique mixtures of languages can be considered to be their own stock. Just as the 6,879 diverse languages of the world each attest to the extent of human cognition, so does each independent contact language. What happens when one of the source languages distinguishes between coarse and refined registers but the other does not, as in the case of Baba Malay (the main component languages of which are Malay and Hokkien)? What happens when a language with clear word

classes is mixed with a language with less clear word classes, as in the case of Kristang (which is predominantly a mixture of Portuguese and Malay)? Is it always a case of simplification, as is expected of contact languages such as creoles (McWhorter 2001)? In fact, no. Baba Malay has invented its own unique coarse-refined register system (Lee 2014) and Kristang has clear word classes (Baxter 1988). What then are all the possibilities when it comes to the creation of new languages in language contact situations? If nothing is done about the endangerment of contact languages, the fact of the matter is that linguists might never know.

The aim of this discussion is to highlight the circumstances of contact languages, the accompanying urgency because of the endangerment circumstances in which many of them are found, and the compelling reasons for why these languages should warrant the attention of linguists. As O'Shannessy (2012:97–98) affirms, “contact languages are no less interesting or important than pre-existing languages”. By highlighting the substantial threat that contact languages face, it is hoped that more researchers will be encouraged to work on these languages, be it in the analysis of the risk patterns of these languages, or in their documentation and conservation.

### References

- Adelaar, Willem F.H. 2014. Endangered language with millions of speakers: Focus on Quechua in Peru. *JournaLIPP* 3. 1–12.
- Ansaldo, Umberto & Lisa Lim. 2014. The lifecycle of Sri Lanka Malay. Special Issue, *Language Documentation & Conservation* 7. 100–118. <http://hdl.handle.net/10125/24605>.
- Baxter, Alan N. 1988. *A grammar of Kristang*. Canberra: Australian National University. Doctoral dissertation.
- Benton, Richard A. 2015. Perfecting the partnership: Revitalising the Mori language in New Zealand education and society 1987–2014. *Language, Culture and Curriculum* 28(2). 99–112.
- Campbell, Lyle, Nala H. Lee, Eve Okura, Sean Simpson & Kaori Ueki. 2013. New knowledge: findings from the *Catalogue of Endangered Languages* (ELCat). Paper presented at the 3rd International Conference on Language Documentation and Conservation, Honolulu, February 28–March 3, 2013. <http://hdl.handle.net/10125/26145>.
- Clements, J. Clancy. 2013. Korlai structure dataset. In Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/contributions/40>. (3 July 2017).
- DeGraff, Michel. 2001. On the origin of creoles: A Cartesian critique of Neo-Darwinian linguistics. *Linguistic Typology* 5(2/3). 213–310.

- DeGraff, Michel. 2017. The sustainable (re)vitalization of local languages is indispensable for education, development and social justice: The MIT-Haiti Initiative as case study. Hong Kong. Workshop session at the 2nd International Conference on Documentary Linguistics–Asian Perspectives: Transformations and Sustainability, May 25–27, 2017.
- Drager, Katie. 2012. Pidgin and Hawai‘i English: An overview. *International Journal of Language, Translation and Intercultural Communication* 1(1). 61–73.
- Erhart, Sabine & Peter Mühlhäusler. 2007. Pidgins and creoles in the Pacific. In Miyaoka, Osahito, Osamu Sakiyama & Michael Krauss (eds.), *The vanishing languages of the Pacific Rim*, 118–143. Oxford: Oxford University Press.
- Evans, Nicholas & Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioural and Brain Sciences* 32. 429–448.
- Farquharson, Joseph T. 2013. Jamaican structure dataset. In Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/contributions/8>. (7 July 2017).
- Fishman, Joshua. 1991. *Reversing language shift*. Clevedon: Multilingual Matters.
- Frank, David B. 2007. We don’t speak a real language: Creoles as misunderstood and endangered languages. Paper presented at the Symposium on Endangered Languages sponsored by the National Museum of Language, March 25, 2007.
- Garrett, Paul B. 2006. Contact languages as “endangered” languages: What is there to lose?. *Journal of Pidgin and Creole Languages* 21(1). 175–190.
- Garrett, Paul B. 2012. Dying young: Pidgins, creoles and other contact languages as endangered languages. In Sodikoff, Genese Marie (ed.), *The anthropology of extinction: Essays on culture and species death*. Bloomington, Indiana: Indiana University Press.
- Girault, Louis. 1989. *Kallawaya: El idioma secreto de los incas*. [Kallawaya: The secret language of the Incas]. La Paz, Bolivia: UNESCO-OPS-OMS.
- Grenoble, Lenore A. & Lindsay J. Whaley (eds.) 1998. *Endangered languages: Language loss and community response*. Cambridge: Cambridge University Press.
- Häberl, Charles G. In press. The Middle East and North Africa. In Moseley, Christopher (ed.), *Encyclopedia of the world’s endangered languages*.
- Hale, Kenneth. 1992. Language endangerment and the human value of linguistic diversity. *Language* 68(1). 35–42.
- Harrison, K. David. 2007. *When languages die: The extinction of the world’s languages and the erosion of human knowledge*. Oxford: Oxford University Press.
- Iseke, Judy. 2013. Negotiating Métis culture in Michif: Disrupting Indigenous language shift. *Decolonization: Indigeneity, Education & Society* 2(2):92–116.
- Krauss, Michael. 1992. The world’s languages in crisis. *Language* 68(1). 4–10.
- Lee, Nala H. 2014. A grammar of Baba Malay with sociophonetic considerations. Honolulu, HI: University of Hawai‘i PhD dissertation.
- Lee, Nala H. 2017. The vitality or endangerment of some nonindigenous languages: a response to Mufwene. *Language* 93(4):e234–e242.

- Lee, Nala H. & John Van Way. 2016. Assessing levels of endangerment in the *Catalogue of Endangered Languages* (ELCat) using the Language Endangerment Index (LEI). *Language in Society* 45(2). 271–292.
- Lehmann, Christian. 1999. Documentation of endangered languages: A priority task for linguistics. *ASSIDUE 1: Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt*, 1–15. Erfurt: Universität Erfurt.
- Lewis, M Paul & Gary F. Simons. 2010. Assessing endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique* 55(2). 103–120.
- Markey, Thomas L. 1982. Afrikaans: Creole or non-creole. *Zeitschrift für Dialektologie und Linguistik* 49.169–207.
- McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2–3). 125–126.
- McConvell, Patrick & Felicity Meakins. 2005. Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics* 25(1). 9–30.
- Meakins, Felicity. 2013. Gurindji Kriol structure dataset. In Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/contributions/72>. (3 July 2017).
- Mühleisen, Susanne. 2013. Trinidad English Creole structure dataset. In Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/contributions/6>. (7 July 2017).
- Mühlhäusler, Peter. 2003. Language endangerment and language revival. *Journal of Sociolinguistics* 7(2). 232–245.
- Mufwene, Salikoko S. 2003. Language endangerment: What have pride and prestige got to do with it? In Joseph, Brian D., Johanna Destafano, Neil G. Jacobs & Ilse Lehiste (eds.) *When languages collide: Perspectives on language conflict, language competition, and language co-existence*. Columbus: Ohio State University Press.
- Mufwene, Salikoko S. 2009. Kitbua, Kileta, Kikongo: What's in a name? In de Féral, Carole (ed.), *Naming languages in Sub-Saharan Africa: Practices, names, categorisations*, 211–222. Louvain-la-Neuve: Peeters.
- Mufwene, Salikoko S. 2017. Language vitality: The weak theoretical underpinnings of what can be an exciting research area. *Language* 93(4). 202–223.
- O'Shannessy, Carmel. 2005. Light Warlpiri: A new language. *Australian Journal of Linguistics* 25(1). 31–57.
- O'Shannessy, Carmel. 2012. Language contact and change in endangered languages. In Austin, Peter K. & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 78–99. Cambridge: Cambridge University Press.
- Sankar, Lokasundari, David Hock Jin Neo & Antoon De Rycker. 2016. Chinese culture and customs in Peranakan funerals in Malaysia and Singapore. *SEARCH: The Journal of the South East Asia Research Centre for Communications and Humanities* 8(1). 17–36.

- Simons, Gary F. & Charles D. Fennig. 2017. *Ethnologue: Languages of the world, 20th edition*. Dallas, Texas: SIL International. <http://www.ethnologue.com>. (13 July 2017).
- van Sluijs, Robbert. 2013. Negerhollands structure dataset. In Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/contributions/27>.
- Smith, Geoff P. & Jeff Siegel. 2013. Tok Pisin structure dataset. In Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/contributions/22>. (7 July 2017).
- Thomason, Sarah G. 1997. A typology of contact languages. In Spears, Arthur K & Donald Winford (eds.), *The structure and status of pidgins and creoles*. Amsterdam: John Benjamins.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Tsunoda, Tasaku. 2005. *Language endangerment and language revitalization: An introduction*. Berlin: Mouton de Gruyter.
- UNESCO ad hoc expert group on endangered languages (Brenzinger, Matthias, Arienne M. Dwyer, Tjeerdde Graaf, Collette Grinevald, Michael Krauss, Osahito Miyaoka, Nicholas Ostler, Osamu Sakiyama, María E. Villalón, Akira Y. Yamamoto & Ofelia Zapeda). 2003. Language vitality and endangerment. Document submitted to the International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages, Paris, 10–12 March 2003. <http://www.unesco.org/culture/ich/doc/src/00120-EN.pdf>. (23 January 2018).
- Valdman, Albert. 1987. Le cycle vital créole et la standardisation du créole haïtien. [The Creole life cycle and the standardization of Haitian Creole]. *Etudes Créoles* 15. 8–94.
- Velayutham, Selvaraj. 2007. *Responding to globalization: Nation, culture, and identity in Singapore*. Singapore: Institute of Southeast Asian Studies.
- Velupillai, Viveka. 2015. *Pidgins, creoles and mixed languages: An introduction*. Amsterdam: John Benjamins.
- Whalen, D. H. & Gary F. Simons. 2012. Endangered language families. *Language* 88(1). 155–173.

Nala H. Lee  
 nala.lee@nus.edu.sg

**Appendix**

Scale of intergenerational transmission

5	4	3	2	I	0
CRITICALLY ENDANGERED	SEVERELY ENDANGERED	ENDANGERED	THREATENED	VULNERABLE	SAFE
There are only a few elderly speakers.	Many of the grandparent generation speak the language, but the younger people generally do not.	Some adults in the community are speakers, but the language is not spoken by children.	Most adults in the community are speakers, but children generally are not.	Most adults and some children are speakers.	All members of the community, including children, speak the language.

Scale of absolute number of speakers

5	4	3	2	I	0
CRITICALLY ENDANGERED	SEVERELY ENDANGERED	ENDANGERED	THREATENED	VULNERABLE	SAFE
1–9 speakers	10–99 speakers	100–999 speakers	1000–9999 speakers	10,000–99,999 speakers	≥ 100,000 speakers

Scale of speaker number trends

5	4	3	2	I	0
CRITICALLY ENDANGERED	SEVERELY ENDANGERED	ENDANGERED	THREATENED	VULNERABLE	SAFE
A small percentage of the community speaks the language, and speaker numbers are decreasing very rapidly.	Less than half of the community speaks the language, and speaker numbers are decreasing at an accelerated pace.	Only about half of community members speak the language. Speaker numbers are decreasing steadily, but not at an accelerated pace.	A majority of community members speak the language. Speaker numbers are gradually decreasing.	Most members of the community speak the language. Speaker numbers may be decreasing, but very slowly.	Almost all community members speak the language, and speaker numbers are stable or increasing.

Scale of domains of use

5	4	3	2	I	0
CRITICALLY ENDANGERED	SEVERELY ENDANGERED	ENDANGERED	THREATENED	VULNERABLE	SAFE
Used only in a few very specific domains, such as in ceremonies, songs, prayer, proverbs, or certain limited domestic activities.	Used mainly just in the home and/or with family, and may not be the primary language even in these domains for many community members.	Used mainly just in the home and/or with family, but remains the primary language of these domains for many community members.	Used in some nonofficial domains along with other languages, and remains the primary language used in the home for many community members.	Used in most domains except for official ones such as government, mass media, education, etc.	Used in most domains, including official ones such as government, mass media, education, etc.