

Chapter 6

Paradigmatic restructuring: The case of Northern Indo-Portuguese Creoles

Clancy Clements

Indiana University

Good (2012) discusses the claim that '[t]he world's most paradigmatically simplified grammars are jargonized grammars.' In this contribution, I consider the process of paradigmatic restructuring resulting from jargonization from the perspective of Klein and Perdue's (2002, 1997) notion of 'basic variety'. Good's concept of 'jargonization bottleneck' is recast as an instance of form selection constrained by frequency, lexical connections, and detectability in the process of naturalistic subsequent language acquisition. In this context, the case of verbal paradigm reduction in the northern Indo-Portuguese creoles is presented. Two recent developments in Korlai are also discussed: the addition of a fourth verb class and the creation of a new paradigm with 'when' conjunctions. This evidence shows that paradigmatic structure can and does again take form.

1 Introduction

Good (2012) advances the claim that creole languages, as a group, are paradigmatically simple but syntactically average. Given that the nature of his paper was, as he expressed it, "largely programmatic", it did not establish a verifiable claim about all contact languages, but did offer a more narrowly focused claim that "[t]he world's most paradigmatically simplified grammars are jargonized grammars" (2012: 37). In other words, his claim is that languages that have historically undergone jargonization will show a reduction in paradigmatic structure that would be detectable typologically. This paper has two goals: first, I will examine what the processes of "jargonization" and concept of "bottleneck" might involve,



and second, I will highlight more recent developments that suggest complexification has already begun in the Indo-Portuguese (IP) creoles examined. For the first goal, I recast the two notions in question within the model developed by (Klein & Perdue 1992, 1997), based on their study of naturalistic second language acquisition and their notion of the “basic variety”. I then consider this approach as it applies to the process of abrupt creolization, which I assume to have taken place in the formation of the IP creoles. For the second, I discuss the addition of a fourth verb class in the IP creoles spoken in Korlai and Daman and the development of a new paradigmatic structure in Korlai IP, which are sensitive to the realis-irrealis distinction.

2 The “basic variety”, jargonization, and the bottleneck

2.1 The “basic variety” and the formation of jargons, pidgins, and creoles

Before the discussion on Klein and Perdue’s notion of the “basic variety”(see discussion below), I will briefly define, as a point of departure, the other terms. A jargon is a rudimentary communication system preceding the development of a pidgin. It may have limited vocabulary, messages such as ‘give me this’, ‘what do you want’, ‘a little bit’ etc. (Bakker 1994: 29), and may also reflect successful attempts at communication in people’s experimentation with forms and structure before systematic conventionalization takes place (Bakker 2003).

The notion of “bottleneck” refers to the result of a kind of language transmission “that is qualitatively distinct not only from ‘canonical’ transmission of language between generations but also from the transmission involved in successful second language acquisition which, while not resulting in full grammatical transmission, does have an end stage wherein the transmitted grammar can be viewed as a variant of the original language” (Good 2012: 4). That is, lexical items may be transmitted but no, or only some, grammatical marking. Good (2012: 6) uses the notion of transmission in a broad sense, “to refer to transmission from any of the languages in a contact situation into an emerging jargon”, without presupposing the speakers fully replicate linguistic material from the source language, into the jargon. I would like to recast these two notions (jargon and bottleneck) within Klein and Perdue’s “basic variety”, to be outlined below, in an attempt to frame the discussion of jargonization within the process of naturalistic language acquisition/creation. It is reasonable to assume that the formation of jargons, pidgins, and creoles involves, very broadly, language acquisition, language processing, and language production, as well as innovation and propagation of linguistic

forms. In the process that leads to the formation of a jargon, and subsequently of a pidgin or a creole, the principle of uniformity (see, for example, McColl Millar 2007: 360–361) allows us to expect that the same laws governing language acquisition in general, including language processing/production, also apply in the conventionalization process that yields a new language variety. With regard to language acquisition proper, the principle would mean that in the formation of a new language variety, the co-creators would process input as second language learners process input: they would acquire content words such as nouns, verbs, and adjectives before function words such as prepositions or verb auxiliaries. They would acquire content words first because these denote entities, activities, events, and states in the world, which tend to be phonetically more substantial than function words and thus more detectable in the speech chain. As a corollary, such learners would also acquire more frequently used forms first before less frequently used forms assuming both are equally detectable. I will return to this below.

Perdue (1993a,b) and Klein and Perdue (1992, 1997) report on findings carried out as part of a project named “Second Language Acquisition by Adult Immigrants” that took place between 1981 and 1988 in five European countries (France, Germany, Great Britain, The Netherlands, and Sweden). In the study, the speech of 40 adult learners representing six native and five target languages was analyzed over a two and one-half year period: native Punjabi and Italian speakers learning English, native Italian and Turkish speakers learning German, native Turkish and Arabic speakers learning Dutch, native Arabic and Spanish speakers learning French, and native Spanish and Finnish speakers learning Swedish. In their data analysis, three developmental stages were identified:

1. Stage 1: Nominal utterance organization (NUO), during which speakers’ utterances contain nominal elements;
2. Stage 2: Infinite utterance organization (IUO), during which speakers’ utterances also contain verbs that are not marked for tense or aspectual distinctions;
3. Stage 3: Finite utterance organization (FUO), during which speakers’ utterances contain nominal and verbal elements and there is evidence of tense/aspect marking.

The Klein and Perdue note that, of the 40 learners whose speech they studied, about two-thirds (27/40) reached the FUO stage, all reached the IUO stage, and

that up to the IUO stage the development of all participants was found to be similar in that they all developed a relatively stable system to express themselves. This system, which was primarily determined by the interaction of a small number of organizational principles, was largely (though not totally) independent of the specifics of source and target language organization. Moreover, it was simple, versatile, and highly efficient for most communicative purposes.

For about one-third of the participants (13/40) in the study, they note that “acquisition ended on this structural level; some minor variation aside, they only increased their lexical repertoire and learnt to make more fluent use of the BV [basic variety]” (Klein & Perdue 1997: 303). Specifically, the learners developed a versatile, functional basic learner variety governed generally by three constraints: the focus of the utterance appears last (pragmatic), the controller of the utterance appears first (semantic), and the most common syntactic orderings in utterances are: NP1 V NP2, NP1 Cop NP2/AdjP/PP, V NP2 (in presentational utterances only).

Assuming the notion of “basic variety” is on the right track, its relation to jargonization and relevance for pidgin/creole formation can be formulated as follows:

- The notion of jargonization discussed in Good (2012) – with subsequent though variable conventionalization of structural patterns in communicative discourse – fits within the NUO-IUO stages, from stage 1 to stage 2 of language development;
- In the NUO and IUO stages, the three general constraints influence the formation of linguistic structures.

Thus, one would expect that, given certain conditions, speakers would not strive to learn target-language systems in order to communicate, but rather would build their own linguistic systems using the material they had at their disposal. This developmental process would be largely governed by the aforementioned basic principles, as well as by (I claim) cognitive processes interacting with frequency of occurrence, including lexical connections, and detectability of elements in the speech chain. If jargons, pidgins, and creoles share comparable acquisitional histories whereby speakers developing them were guided by basic constraints in situations of continuous communicative interaction, material that is superfluous for communication, the verbal paradigms encoding tense, mood, aspect, person, and number, for example, would not be acquired. Rather, speakers would create other means using the lexical items they have acquired to express

what they need. The question about what lexical items speakers acquire in such situations involves the notion of “bottleneck”, which will now be discussed in further detail.

2.2 The “bottleneck” and the formation of jargons, pidgins, and creoles

As stated above, the notion “bottleneck” refers to the result of a kind of language transmission in the language acquisition process. It does not necessarily result in full grammar transmission of a target language, but can result in a system where a speaker’s target language grammar is a variant of that of the target language. Within a contact situation, Good (2012: 6) states that transmission can happen from any of the languages in contact into an emerging jargon. As a working hypothesis, I assume that the conventionalization process would likely follow the organization principles identified by Klein & Perdue (1992, 1997) and discussed above.

2.2.1 Frequency

In the acquisition process, the frequency with which a lexical form appears in discourse impacts when in the acquisition process it may be acquired. To capture this, (Andersen & Shirai 1996) proposed the Distribution Bias Hypothesis (DBH) to account for which forms of a lexical item are acquired first in the acquisition process. In essence, the DBH states that in L2 acquisition the frequency with which, say, a verb form appears in language use affects the order in which it is first acquired.¹

To give an illustrative example, in the data I have on Chinese Immigrant Spanish, one of the speakers talked about a number of events related to their background. The verb form used in the narrative is overwhelmingly the 3sg present-tense form, independently of person, number, or temporal/aspectual reference. This is the case, as will be discussed in more detail below, because of the conjugated forms of any given verb the 3sg present-tense form is the most frequently

¹Related to the DBH is the Primacy of Aspect Hypothesis (POA) (Andersen 1993, Andersen & Shirai 1996). The POA states that the lexical aspect of a verb is important in determining in which verb forms a given verb more commonly appears. For example, stative verbs (such as Spanish *saber* ‘know’) and atelic dynamic verbs (such as *andar* ‘walk’) which appear more commonly in imperfective forms, than do telic punctual verbs (such as *nacer*, *nacer* ‘be born’ or *chegar*, *llegar* ‘arrive’), which appear more commonly in perfective forms than do atelic dynamic and stative verbs.

occurring in written and spoken Portuguese and Spanish, as shown by the token counts of verb forms from the Davies (2002) language corpora data. There is an exception to this strong pattern though. When relating differences about their parents' backgrounds (place of birth), the speaker in question used the 3SG preterit form *nació* 'was born' (a telic punctual verb). The immediate question arises: why wasn't the 3SG present-tense form selected? Again, the Davies' Spanish-language corpus provides a reasonable answer if we assume the DBH, as well as the POA. Of all the present-tense, preterit, imperfect, and infinitive forms of *nacer* found in the corpus, the 3SG preterit form is the most frequent one, with 15% more tokens than the next most frequent form (See Table 1).

Table 1: Present-tense, preterit, imperfect, and infinitival forms of *nacer* 'to be born' in Davies' Corpus del Español (2002), listed by token frequency.

| P/N | T/A | Form | Token | % |
|----------|-------|-----------|-------|-----|
| 3SG | PRET | nació | 1,643 | 35 |
| 3SG | PRS | nace | 975 | 20 |
| | INF | nacer | 802 | 17 |
| 3PL | PRS | nacen | 568 | 12 |
| 3PL | PRET | nacieron | 300 | 6 |
| 1SG, 3SG | IMPFT | nacía | 154 | 3 |
| 3PL | IMPFT | nacían | 84 | 2 |
| 2SG | PRET | naciste | 80 | 2 |
| 1PL | PRS | nacemos | 44 | 1 |
| 1PL | PRET | nacimos | 42 | 1 |
| 2SG | PRS | naces | 14 | 0.3 |
| 2PL | PRET | nacisteis | 8 | 0.2 |
| 1SG | PRS | nazco | 6 | 0.2 |
| 2PL | PRS | nacéis | 1 | 0.1 |
| 2SG | IMPFT | nacías | 1 | 0.1 |
| 1PL | IMPFT | nacíamos | 1 | 0.1 |
| 2PL | IMPFT | nacíais | 0 | 0 |
| Total | | | 4,723 | 100 |

Although the use of a corpus in this manner is admittedly an imperfect tool and only an approximate manner of gauging frequency of use of verb forms of a

paradigm in discourse, it nevertheless makes the correct prediction in the case of the forms of *nacer*, as well as in the large majority of cases, as we will see below.

Returning to how to recast the process of jargonization, then, the relative frequency of a form in discourse, and I will argue its detectability and lexical connections as well, are important in how speakers build their grammars.

If we assume that in the formation of an immigrant variety (as in the discussion of the “basic variety”), a jargon, a pidgin, or a creole, speakers are negotiating a system of communication in which lexical meaning is most important and grammatical meaning is deduced through the situation and the context (i.e. pragmatically, as suggested by Mühlhäusler 1997), speakers will acquire word forms that are most frequently used in discourse, most lexically connected, and most easily detectable. We can define frequency as the number of times in a given corpus that a certain item or form appears. Following the aforementioned principle of uniformity, we assume as a working hypothesis that the most frequently used word forms in corpora are, roughly speaking, also the most frequently used forms in discourse in a contact situation and the ones that would be the most likely candidates for selection in the formation of a jargon, pidgin, or creole.

The data I consider in this section suggests that, in the process of jargonization, the selection of one form over others can be accounted for by appealing to frequency of occurrence, already touched upon; lexical connectedness of forms that are likely candidates for selection in the conventionalization process; and the detectability of forms that are likely candidates for selection in the conventionalization process.

To give an example, in the Portuguese-language corpus of Davies & Ferreira (2006), the most frequently occurring present-tense form of the atelic dynamic verb *beber* ‘drink’ is the 3SG form *bebe*, while the most frequently occurring non-finite form by far is the infinitive form *beber* (see Table 2).

Thus, in a situation in which an immigrant variety, pidgin, or creole is taking shape, if frequency of occurrence plays a role in the form selection process, and if learners were to select from among the candidate forms listed in Table 2, they would select *beber*, the overall most frequently occurring form, and after that *bebe*, the second-most frequently occurring form. As we will see below, the base form in the contact varieties studied in this paper derive either from the infinitival form, or from the 3SG present-tense form, the two most frequently occurring forms in the Davies & Ferreira (2006) data base.²

Bybee (1985: 123–27) discusses another factor linked to frequency, which she calls lexical connections; that is, shared phonetic material among different forms.

²For purposes of exposition, the non-present forms are not included, as they are numerically weakly represented relative to the present-tense and infinitival forms.

Table 2: Lexical frequency (strength) of present-tense and some non-finite forms of *beber* ‘drink’ from two periods (Davies and Ferreira 2006).

| P/N | Form | 1500–1799 | | 1900s | |
|-------|---------|-----------|-----|----------|-----|
| | | <i>n</i> | % | <i>n</i> | % |
| INF | beber | 682 | 73 | 580 | 68 |
| 1SG | bebo | 9 | 0.8 | 42 | 5.5 |
| 2SG | bebes | 6 | 0.5 | 28 | 3 |
| 3SG | bebe | 155 | 17 | 128 | 15 |
| 1PL | bebemos | 8 | 0.7 | 21 | 3 |
| 2PL | bebeis | 0 | 0 | 0 | 0 |
| 3PL | bebem | 73 | 8 | 46 | 5.5 |
| Total | | 933 | 100 | 845 | 100 |

In a paradigm, such as the forms in Table 2, there are two key sets of lexical connections, shown in Figure 1. These are distinguished by stress assignment: In the left hand column, the connection common to all forms is *‘bebe*, indicated by the vertical lines.

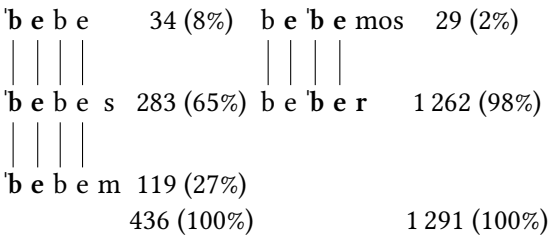


Figure 1: Lexical connections of present-tense and the infinitive forms of *beber* ‘drink’ (both periods combined).

In the righthand column, the connection common to all forms is *be’be*, also indicated by the vertical lines. If lexical connections are important for form selection in naturalistic second (or subsequent) language acquisition in a language contact situation, the most strongly represented candidate, in terms of frequency and lexical connections, would be *be’be* and it would also be the preferred candidate over *‘bebe*, the second most-preferred candidate. The same is applicable to Spanish. Assuming, then, that the relative distribution of forms of *beber* is

comparable for other atelic dynamic verbs, such as *cantar* ‘sing’ or *andar* ‘walk’, it is possible to extrapolate, and advance the claim that, in the conventionalization process of a language contact variety such as an immigrant variety, a pidgin, or a creole, there are two main candidates for selection in the present-tense Portuguese and Spanish paradigms, $\sigma\sigma$ and $\sigma'\sigma$, and the latter will be selected. As indicated, these predictions are largely borne out in the data sample used in Clements (2014, 2018). The results are given in Table 3. We see, then, that the infinitival form of the lexifier language is most often preferred, and if there is a secondary form, it is the 3SG form. Thus, taking lexical strength and lexical connections into account allows us to formulate a falsifiable hypothesis about form selection in the conventionalization process of contact language such as an immigrant variety, a pidgin, or a creole.

Table 3: Default form and verb form source

| Language | Verb form source from Portuguese/Spanish |
|------------------------------|--|
| a. Angolar | INF |
| b. Papiamentu | INF and 3SG |
| c. Palenquero | INF |
| d. Bozal Spanish | INF and 3SG |
| e. Chinese Coolie Spanish | variable but 3SG (62%) and INF (17.5%) preferred |
| f. Macau Portuguese | INF |
| g. Chinese Immigrant Spanish | variable but 3SG (48%) and INF (19%) preferred |
| h. Korlai Indo-Portuguese | INF |
| i. Daman Indo-Portuguese | INF |

2.2.2 Detectability

The notion of detectability (see 1 below), is based on two uncontroversial distinctions and one descriptive observation. The observation is the ubiquity of CV structure in the world’s spoken languages. As for the distinctions, those between stress-bearing vs. non-stress-bearing syllables between free vs. bound morphemes are relevant here. Thus, for the purposes at hand detectability is defined in relative terms: syllables containing or consisting of a CV structure are more easily detected in the speech chain than those without a CV structure (e.g.

V or VC structure), and that stressed syllables and free-standing morphemes are more easily detected in the speech chain than unstressed syllables and clitics/affixes, respectively. This can be stated as in (1).

- (1) Definition of Detectability
 - a. CV is more detectable than VC, V
 - b. stressed syllables are more detectable than unstressed syllables
 - c. free-standing morphemes are more detectable than clitics and affixes

Based on the foregoing, then, I assume that in the conventionalization process of a language-contact variety, the nature and extent of restructuring, and thus the relative importance of frequency, lexical connections, and detectability in shaping a newly-emerging language variety, depend on the individual makeup of a given contact situation. At the same time, it must be acknowledged that the adult agents of the acquisition/creation process already know one or more languages and that in creating a language contact variety, speakers may introduce into the new language elements from their own language(s). In the literature, this is variably called imposition or interference through shift.

Having defined frequency, lexical connections, and detectability, I would like to illustrate now how they operate in the selection process of a form, namely, the copula, in the aforementioned contact varieties. It turns out that frequency and detectability seem to interact in the selection process.

2.2.3 Illustrative example of the “bottleneck”: Frequency, lexical connections, and detectability in copula selection

In Portuguese and Spanish, there are two copulas: *ser* and *estar*. Restricting myself to the infinitive and present-tense forms for the present discussion (these are the most relevant), I assume that all forms of both copulas were possible candidates for selection in the formation of the contact varieties in question. Based on frequency counts gleaned from the Davies & Ferreira (2006) Portuguese-language corpus and the Davies (2002) Spanish-language corpus, shown in Tables 4 and 5, and Tables 6 and 7, we see that the 3SG is by far the most frequently occurring form in both languages, for all periods (16th–18th c., 18th–19th c.) and in both genres (written vs. oral). The second most frequently occurring form is different for the two languages. In Portuguese, the infinitival form *ser* is more frequent than *são* 3PL in the written data (16th to 20th century), but in the oral data the two forms are roughly equally frequent. In Spanish, while *ser* is more

frequent in the 16th–18th centuries, in the 19th–20th centuries the two forms *ser* and *son* 3PL are roughly equally frequent and in the oral data *son* is more frequent. Thus, it is reasonable to say that, in terms of frequency, these two forms competed with one another in the selection process. Given that the other forms of the respective paradigms are rather weakly represented, based on frequency alone, that is, the frequency distributions in Tables 4 and 5, the following predictions can be made. If frequency alone is favored in copula selection during the conventionalization process of a contact variety such as an immigrant variety, a pidgin, or a creole, the 3SG copula form *é* for Portuguese and *es* for Spanish should be selected, because they are by far the most frequently occurring forms, respectively. As we will see below, the forms *é*, *es* are seldomly selected. The conclusion is, then, that frequency alone is not a reliable predictor for form selection in the formation of the contact languages being examined here.

Table 4: Distribution of the infinitive and present-tense forms of Portuguese *ser* ‘be’ in two periods and two genres

| Form | 16th–18th c. | | 19th–20th c. | | Oral | |
|------------------|--------------|---------|--------------|--------|--------|---------|
| 3SG <i>é</i> | 21,261 | (41.5%) | 249,188 | (52%) | 47,681 | (78.6%) |
| INF <i>ser</i> | 18,777 | (37%) | 161,781 | (33%) | 5,808 | (9.6%) |
| 3 PL <i>são</i> | 9,294 | (18%) | 59,285 | (12%) | 5628 | (9.3%) |
| 1SG <i>sou</i> | 1,132 | (2%) | 9,109 | (2%) | 1,153 | (2%) |
| 1PL <i>somos</i> | 487 | (1%) | 2,060 | (0.4%) | 335 | (.45%) |
| 2SG <i>és</i> | 207 | (0.5%) | 2,855 | (0.6%) | 34 | (.05%) |
| Total | 51,158 | (100%) | 484,278 | (100%) | 60,639 | (100%) |

With respect to the other copulas, forms of Portuguese and Spanish *estar* ‘to be’, the frequency distribution of the same forms (infinitival and present-tense) are shown in Tables 6–7.

The most frequently occurring form by far in both languages, in all centuries and in both genres, is the 3SG form *está*. Thus, in terms of frequency of occurrence, this form would be the main competitor to *é*, *es*. But there are other factors that we need to take into consideration, namely lexical connections and detectability. As stated above, lexical connections refers to shared phonetic material among different forms. The lexical connections of the copula forms are displayed in Tables 8 and 9.

In terms of detectability, the lexical connections involving Portuguese *é*, *és* and Spanish *és*, *eres* in Table 8 are disfavored because the forms either do not

Table 5: Distribution of the infinitive and present-tense forms of Spanish *ser* 'be' in two periods and two genres

| Form | 16th–18th c. | | 19th–20th c. | | Oral | |
|-----------|--------------|--------|--------------|--------|--------|--------|
| 3SG es | 284,568 | (62%) | 257,156 | (67%) | 66,804 | (77%) |
| INF ser | 81,376 | (18%) | 55,240 | (14%) | 7,351 | (8%) |
| 3PL son | 67,371 | (14%) | 54,643 | (14%) | 10,239 | (12%) |
| 1SG soy | 17,548 | (4%) | 10,210 | (3.3%) | 1,569 | (1.5%) |
| 2 SG eres | 6,768 | (1.5%) | 4,019 | (1%) | 337 | (.5%) |
| 1PL somos | 2,445 | (0.5%) | 3,113 | (0.7%) | 894 | (1%) |
| Total | 460,076 | (100%) | 384,381 | (100%) | 87,194 | (100%) |

Table 6: Distribution of the infinitive and present-tense forms of Portuguese *estar* 'be' in two periods and two genres.

| Form | 16th–18th c. | | 19th–20th c. | | Oral | |
|-------------|--------------|---------|--------------|--------|--------|--------|
| 3SG está | 4,852 | (43.5%) | 42,105 | (53%) | 6,666 | (54%) |
| INF estar | 3,066 | (27.5%) | 8,736 | (11%) | 1,174 | (9.5%) |
| 3 PL estão | 1,723 | (15.5%) | 13,795 | (17%) | 1,888 | (15%) |
| 1SG estou | 984 | (9%) | 9,325 | (11%) | 1,445 | (12%) |
| 1PL estamos | 412 | (3.5%) | 3,792 | (5%) | 1,061 | (9%) |
| 2SG estás | 114 | (1%) | 1,950 | (3%) | 54 | (0.5%) |
| Total | 11,151 | (100%) | 79,703 | (100%) | 12,288 | (100%) |

Table 7: Distribution of the infinitive and present-tense forms of Portuguese *estar* 'be' in two periods and two genres.

| Form | 16th–18th c. | | 19th–20th c. | | Oral | |
|-------------|--------------|---------|--------------|--------|--------|--------|
| 3SG está | 45,483 | (49%) | 47,014 | (49%) | 12,251 | (47%) |
| 3PL están | 16,135 | (17.5%) | 18,311 | (19%) | 4,614 | (18%) |
| INF estar | 13,481 | (15%) | 10,581 | (11%) | 2,257 | (9%) |
| 1SG estoy | 11,089 | (12%) | 9,944 | (10%) | 2,760 | (11%) |
| 1PL estamos | 2,242 | (2.5%) | 6,500 | (7%) | 3,192 | (12%) |
| 2SG estás | 3,727 | (4%) | 3,523 | (4%) | 856 | (3%) |
| Total | | (100%) | 95,873 | (100%) | | (100%) |

Table 8: Lexical connections of the infinitive and present-tense forms of Portuguese and Spanish *ser*

| Portuguese | | Spanish | |
|------------|-------|-----------|---------|
| CV(C) | V(C) | CV(C) | V(C) |
| s o u | é s | s o y | e r e s |
| | | | |
| s o m o s | é | s o m o s | e (s) |
| | | | |
| (s o i s) | | | |
| \ | | | |
| s ã o | s o n | | |
| | | | |
| s e r | s e r | | |

Table 9: Lexical connections of the infinitive and present-tense forms of Portuguese and Spanish *estar*

| Portuguese CC-'CV(C/G) | Spanish CC-'CV(C) |
|---------------------------|----------------------|
| e s t o u | e s t o y |
| | |
| e s t á s | e s t á s |
| | |
| e s t á | e s t á |
| | |
| e s t a m o s | e s t a m o s |
| | |
| e s t ã o | e s t á n |
| | |
| e s t a r | e s t a r |

contain or consist of a CV structure, or are infrequent (*és, eres*), or both. Thus, it would be predicted that these forms would not be selected. By contrast, the other sets of lexical connections in Tables 8 and 9 do include frequently occurring forms (Portuguese *são, está*; Spanish *son, está*) that contain a CV structure. Based on this observation, two predictions can be advanced. If detectability alone is favored in copula selection, we should find a number of candidates selected that consist of, or contain, a CV structure. These candidates are Portuguese *sou, estou, estás, está, estar, somos, estamos, ser, são, estão* and Spanish *soy, estoy, estás, está, somos, estamos, son, ser, están*. If the combination of detectability and frequency is favored in copula selection, the prediction is that we should find few forms (or reflexes thereof) as copulas, namely, Portuguese *são, ser, está* and Spanish *son, ser, está*. If lexical connections is also taken into consideration, the infinitival form *ser* in both Portuguese and Spanish is disfavored, but *estar* adds its shared phonetic material to the most frequently occurring form *está*. The combination of frequency, detectability, and lexical connections in form selection turns out to be important in the great majority of cases. The copulas of the varieties examined here are shown in Table 10. Note that in almost all cases the predicted forms, that is, reflexes of *são/son, está(n/r)/estão* were selected.

In those cases in which a form was selected outside of the candidate pool shown in Tables 8 and 9, some of the aforementioned predictions would still apply. That is, any form selected as a copula would have to be one of the most frequently occurring forms of its paradigm and contain, or consist of, a CV structure. This prediction is borne out in all relevant cases except one, that of *sendá* in Palenquero (in Table 10c), whose source is most likely the verb form *sentar* [sen-‘tar] ‘sit down’. All other forms are indeed 3SG forms that contain a CV structure: Palenquero *fwe* (in Table 10c) and Indo-Portuguese *te, tē* (in Table 10h, Table 10i).³

In three cases, the current model advanced above does not make the correct predictions based on a selective process involving frequency, detectability, and lexical connections. Specifically, the model would not predict Portuguese *é* or Spanish *es* to be selected because they lack a CV structure, although they are by far the most frequently occurring copula forms in the corpora, independently of historical period or genre. There may be other reasons that have influenced the selection process. In the case of Palenquero (in Table 10c), its speakers have been bilingual in Spanish and Palenquero for more than a century. In the second

³Information sources for these contact varieties are Lorenzino 2007 (Angolar), Kouwenberg & Ramos-Michel 2007 and Maurer 1998 (Papiamentu), Schwegler 1998 and Schwegler & Green 2007 (Palenquero), Clements 2009 (Bozal Spanish, Chinese Coolie Spanish, Chinese Immigrant Spanish), Batalha 1974 and Pinharanda 2010 (Macau), Clements 1996 (Korlai), Clements & Koontz-Garboden 2002 (Daman).

Table 10: Contact-language copulas with Portuguese and Spanish source forms. †: infrequent

| Language | Copula | Source form |
|-------------------------------|--------------|--|
| a. Angolar | <i>θa</i> | < Ptg. <i>est-á/-ão/-r</i> |
| | <i>ta</i> | < Ptg. <i>est-á/-ão/-r</i>) |
| b. Papiamentu | <i>ta</i> | < Ptg. <i>est-á/-ão/-r</i>) |
| c. Palenquero | <i>fwe</i> | < Sp. <i>fue</i> ‘was’ |
| | <i>sendá</i> | < Sp. <i>sentar</i> ‘sit down-INF’ |
| | <i>é</i> | < Ptg. <i>é</i> and/or Sp. <i>e(s)</i> |
| d. Bozal Spanish (Cuba) | <i>son</i> | < Sp. <i>son</i> |
| e. Chinese Coolie Span (Cuba) | <i>son</i> | < Sp. <i>son</i> |
| f. Macau Portuguese | <i>sã</i> | < Ptg. <i>são</i> |
| g. Chinese Immigrant Spanish | <i>son</i> | < Sp. <i>son</i> |
| | <i>está</i> | < Sp. <i>est-á/-án/-ar</i> |
| | <i>es</i> † | < Sp. <i>es</i> |
| h. Korlai Indo-Portuguese | <i>tɛ</i> | < Ptg. <i>tem</i> (3SG), <i>têm</i> (3PL) ‘have’ |
| i. Daman Indo-Portuguese | <i>te</i> | < Ptg. <i>tem</i> (3SG), <i>têm</i> (3PL) ‘have’ |
| | <i>é</i> | < Ptg. <i>é</i> |

case, speakers of Daman Indo-Portuguese have had the presence of European Portuguese from the time of its formation in the late 16th century up until 1961, with a reduced presence from that year till the present day. In the third case, that of Chinese Immigrant Spanish, the form *es* is infrequently used in favor of *está* and *son*. Thus, in all three cases there are circumstances that appear to have overridden the prediction that the selected form is predictable from the combination of frequency, detectability, and lexical connections.

Thus, the “bottleneck” referred to in Good (2012) can be recast as a set of restrictions on processability: the relative lack of frequency, detectability, and lexical connections of any given form used in communicative interaction represents a “bottleneck” in the conventionalization process of a language variety that yields an immigrant variety (Klein and Perdue’s “basic variety”), which in turn can give way to a jargon, pidgin, or creole. Stated a different way, the most frequently occurring forms that are most detectable and lexically connected are most likely to be selected as part of the jargonization process. For the nine language varieties being discussed here, the combination of frequency, detectability, and lexical connections gives the correct predictions to a large extent. If frequency alone is considered, that is, the most frequently occurring forms in the corpora consulted, we

only have one part of the bottleneck. If detectability and lexical connections are combined with frequency, the reduced pool of copula candidates (Portuguese *está*, *são* and Spanish *está*, *son*) are the most likely forms to be selected. This statement also applies to Portuguese 3SG *tem* ‘has’, by far the most frequently occurring form of the present-tense paradigm in all historical periods, as well as in the written and oral genres, as shown in Table 11.

Table 11: Frequency counts of Portuguese present-tense forms of *ter* ‘have’ from the 16th to the 20th century, listed in order of frequency, representing written and oral (only 20th c.) genres

| Form | 16th–18th c. | 19th–20th c. | Oral |
|------------------|---------------|---------------|---------------|
| 3SG <i>tem</i> | 17,620 (71%) | 46,740 (57%) | 10,147 (59%) |
| 1SG <i>tenho</i> | 3,433 (14%) | 13,600 (17%) | 3,031 (18%) |
| 3PL <i>têm</i> | 1,638 (7%) | 11,979 (14%) | 2,034 (12%) |
| 1SG <i>temos</i> | 1,813 (7%) | 6,523 (8%) | 1,822 (10.5)2 |
| SG <i>tens</i> | 222 (1%) | 2,798 (4%) | 98 (.5%) |
| Total | 24,726 (100%) | 81,640 (100%) | 17,132 (100%) |

Thus, in the creation of contact varieties such as immigrant language, pidgins, and creoles, the data considered from the nine contact varieties in question strongly suggest that the bottleneck is profitably defined in terms of frequency of the forms in discourse, how much phonetic material is shared by forms in the paradigm (lexical connections) and the syllabic structure of such forms (detectability).

Although the languages being discussed are paradigmatically more streamlined than their respective lexifier languages, they of course have continued to evolve. Interestingly and not unexpectedly, they have developed new paradigmatic structures. I would just like to highlight two such developments in order to suggest that paradigmatic structure in a language may decrease or increase, depending on the nature of the contact situation.

3 Two developments in Korlai IP after its formation

As shown in Appendix A, the northern Indo-Portuguese creoles spoken in Korlai, Daman, and Diu are unique in that they have retained the three verb classes from Portuguese (*-a*, *-e*, *-i*) with the corresponding allomorphy. This suggests that

these creoles have a shared history that goes back to the 16th century. It is noteworthy that no other Portuguese-lexified creoles have retained the morphology present in these Indo-Portuguese creoles, although it is clear the paradigmatic richness is greatly reduced (see Appendix B for comparison). In independent developments, Korlai IP, and to a lesser extent Daman IP, have added a fourth verb class. Moreover, Korlai IP has developed a paradigm involving conjunctions referring to ‘when’. I will briefly describe these two innovations.

3.1 The fourth verb class in Korlai IP

Given that there has not been any systematic investigation of the fourth verb class in Daman IP, my focus here will be on Korlai IP’s fourth verb class. In addition to the three conjugation classes surveyed above, Korlai IP has developed a wholly new conjugation class as a result of the contact between Korlai IP and Marathi: the *-u* class. Korlai speakers use this verb class extensively to borrow verbs from Marathi. The Marathi form is the invariable imperative verb form (see Clements & Luís 2014 for details). The form borrowed in Daman IP is still being investigated. Its verbs are conjugated the same way as verbs belonging to the other classes. The full paradigm *-u* verb class member *loṭú* ‘push’ is shown in Table 12, together with verbs from the other three classes together with verbs from the other three classes (*katá* ‘sing’, *bebé* ‘drink’, and *irgí* ‘get up’).

Table 12: The four verb classes in Korlai IP with illustrative examples.

| | Class 1 | Class 2 | Class 3 | Class 4 |
|---------------|----------------|----------------|----------------|----------------|
| a. Unmarked | <i>kat-á</i> | <i>beb-é</i> | <i>irg-í</i> | <i>loṭ-ú</i> |
| b. Past | <i>kat-ó</i> | <i>beb-é-w</i> | <i>irg-í-w</i> | <i>loṭ-ú</i> |
| c. Gerund | <i>kat-á-n</i> | <i>beb-é-n</i> | <i>irg-í-n</i> | <i>loṭ-ú-n</i> |
| d. Completive | <i>kat-á-d</i> | <i>beb-í-d</i> | <i>irg-í-d</i> | <i>loṭ-ú-d</i> |

The verb forms in Table 12 reveal that *loṭ-ú* ‘push’ takes the same inflectional endings as the verbs in the other three verb classes, with one exception: the past form is syncretic with its unmarked form. All borrowed verbs from Marathi in Korlai IP belong to Class 4 (Clements & Luís 2014).

Semantically, borrowed verbs in general can make different semantic contributions to its lexicon. They may: a) fill in a semantic gap (in which case they constitute an extension to the lexicon), b) be synonymous to already existing verbs, or c) simply replace a word that has become obsolete (Wohlgemuth 2009).

In the case of Korlai, some loan verbs are synonyms of native Korlai verbs while others are not. An example of a loan verb and native-verb pair (i.e., lexical variants) is Korlai *av du* ‘like’ (< Marathi *av du* ‘like-IMPERATIVE’) and *gostá* (< Ptg. *gostar* ‘like’), also found as *gostí* ‘ditto’, the form favored by the younger speakers. However, while some loan verbs in Korlai overlap semantically with native Korlai verbs, they do have different contexts of use. Two examples of this are shown in (2).

- (2) Loan Native
 aṭu *finhika* ‘shrink/become small’
 bānu *kudzinya* ‘prepare, cook’

The loan verb *aṭu* ‘shrink’ is used only for clothes and other fabrics that can shrink, and the loan verb *bānu* ‘prepare’ is often used in the context of cooking, but is coming to be used in other contexts, as well, whereas the native verb *kudzinya* ‘cook’ is restricted to food preparation.

3.2 The emergence of a paradigm involving Korlai IP conjunctions ‘when’

The full scope of how ‘when’ is expressed in Korlai IP includes interrogative (direct and indirect) and assertive speech with subordinate clauses headed by conjunctions (finite clauses), as well as postpositions (non-finite clauses). The expression of ‘when’ in subordinate clauses is sensitive to mood (realis-irrealis). That is, mood is coded in the ‘when’ conjunctions in Korlai.

To express direct and indirect interrogative ‘when’ in Korlai, *kər* is used, as shown in (3a) and (3b), respectively.

- (3) a. *Teru kər lə vi?*
 Teru when FUT come
 ‘When will Teru come?’
 b. *Teru kər lə vi, kē sab.*
 Teru when FUT come who know
 ‘Who knows when will Teru come.’

‘When’ in subordinate clauses with future reference can be expressed with the postposition *-ni* in non-finite clauses, as in (4a), or with the conjunction *kərki* in finite clauses, as in (4b).

- (4) a. *Teru vin-ni, nɔ lə anda Boməy.*
 Teru come-when FUT come who know
 ‘When Teru comes, we will go to Mumbai.’
 b. *Teru kərki lə vi, nɔ lə anda Boməy.*
 Teru when FUT come 1PL FUT go Mumbai
 ‘When Teru comes, we will go to Mumbai.’

In subordinate clauses with past-reference ‘when’, the postpositional structure is not used. Rather, the conjunction *ki* expresses ‘when’, as shown in (5). Thus, the use of one or another conjunction depends on mood. If the temporal subordinate clause headed by ‘when’ has past reference, *ki* is used; if it has future reference, *kərki* is used.

- (5) *Teru ki yav e, nɔ yaho Boməy*
 Teru when came 1PL went Mumbai
 ‘When Teru came, we went to Mumbai.’

The conjunction *ki* ‘when’ never appears as the head of a clause with future reference (irrealis), as shown in (6b). Similarly, *kərki* ‘when’ heading a clause with past reference (realis) is equally unacceptable, as shown in (6a).

- (6) a. **Teru kədz kərki jav e nɔ ti kumen.*
 Teru house when came 1PL PST eating
 ‘When Teru came, we were eating.’
 b. **Teru kədz ki lə vi nɔ lə kume.*
 Teru house when FUT come 1PL FUT eat
 ‘When Teru comes, we will eat.’

Thus, what has developed in Korlai I consider to be a two-member paradigm with the conjunctions expressing ‘when’ that are sensitive to mood. This is shown schematically in (7).

- (7) [CP XP *ki* [IP _(realis)_]] Matrix Cl
 [CP XP *kərki* [IP _(irrealis)_]] Matrix Cl

These developments in Korlai have taken place more recently though it is impossible to say when. What the developments suggest is that as languages with paradigmatic structures evolve, speakers are likely to create new paradigms that encode distinctions in novel ways using the means available to them.

In sum, in this section I have briefly highlighted two innovations found in Korlai: the development of an additional verb class that Korlai speakers use productively to borrow verbs from Marathi, and the emergence of two-member paradigm consisting of ‘when’ conjunctions whose use are sensitive to the realis-irrealis mood distinction. Interestingly, this distinction found in the selection of one or another ‘when’ conjunction in Korlai encodes part of what Portuguese encodes with different verb forms. In Portuguese, subordinate clauses headed by ‘when’ with past reference contain a preterit (i.e., a realis) verb form (8a), while subordinate clauses headed by ‘when’ with future reference contain a future imperfect (i.e., irrealis) verb form, as in (8b).

- (8) a. Quando chegou Teru, nós fomos para Lisboa.
 when arrive.PST.3SG Teru 1PL go.PST.1PL for Lisbon
 ‘When Teru arrived, we went to Lisbon.’
 b. Quando chegar a Teru, nós vamos para Lisboa.
 when arrive.FUT.IPFV Teru 1PL go.PRS.1PL for Lisbon
 ‘When Teru arrives, we will go to Lisbon.’

4 Concluding remarks

In this contribution, I offered a recast of the process of jargonization and the concept of the bottleneck discussed in Good (2012) as a cognitive process of form selection, within Klein and Perdue’s model of naturalistic L2 acquisition resulting in what they call the “basic variety”. In this model, jargonization can be understood as a speaker’s developmental utterance organization evolving from a stage of nominal utterance organization to a stage of non-finite utterance organization. At this latter stage, the verbal communication system is (according to Klein & Perdue 1992, 1997) efficient for communicative purposes. It is at this stage 2, I have argued, that speakers begin to build their own grammar from the material they already have available to them if the circumstances prompt them to do so. I have proposed that the notion of bottleneck can be understood to refer to a number of cognitive and social aspects of how humans communicate (form selections in parsing and producing utterances) in situations of regular communicative interaction among speakers who do not share a common language. For communication, humans target and acquire frequently used, detectable forms with robust lexical connections and they build their grammar using these forms, again, if the circumstances favor it.

Applying the notions of the basic variety development and form selection, the forms found in the nine contact varieties examined in this contribution can be accounted for in a principled way. In the same way, loss of paradigmatic structure can also be accounted for. Good's claim that creoles, as well as other contact varieties discussed in this paper, are paradigmatically simple can be more comprehensively understood within the model of the basic variety (with its constraints), frequency, detectability, and lexical connections.

This, of course, in no way precludes the development of new categories and new paradigms in such restructured languages and I have discussed two such cases here: the addition of a new verb class in Korlai that is used to accommodate verbs borrowed from Marathi into Korlai, and the emergence of a two-member paradigm in Korlai that contains two 'when' conjunctions, one that encode realis mood (*ki*) found in subordinate 'when' clauses with past reference, and another that encodes irrealis mood (*kərki*) found in subordinate 'when' clauses with future reference. What these developments suggest is that a language variety, even if highly restructured historically, does develop new, more complex, structures and patterns as it evolves, in order to accommodate the needs of its speakers.

Abbreviations

| | | | | | |
|-------|------------|------|--------------|-----|----------|
| FUT | future | IPFV | imperfective | PRS | present |
| IMPFT | imperfect | PL | plural | PST | past |
| INF | infinitive | PRET | preterit | SG | singular |

References

- Andersen, Roger W. 1993. Four operating principles and input distribution as explanations for under-developed and mature morphological systems. In K Hyltenstam & A Viborg (eds.), *Progression and regression in language*, 309–339. Cambridge University Press.
- Andersen, Roger W. & Yasuhiro Shirai. 1996. The primacy of aspect in first and second-language acquisition: The pidgin-creole connection. In William C. Ritchie and Tej K. Bhatia (ed.), *Handbook of second language acquisition*, 527–570. San Diego, CA: Academic Press.
- Bakker, Peter. 1994. Pidgins. In Jacques Arends, Pieter Muysken & Norval Smith (eds.), *Pidgins and creoles: An introduction*, 25–39. Amsterdam: Benjamins.
- Bakker, Peter. 2003. Pidgin inflectional morphology and its implications for creole morphology. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology*, 3–33. Dordrecht: Kluwer.

- Batalha, Graciete Nogueira. 1974. *Língua de Macau*. Macau: Imprensa Nacional de Macau.
- Bybee, Joan. 1985. *Morphology*. Amsterdam & Philadelphia: John Benjamins.
- Cardoso, Hugo C. 2009. *The Indo-Portuguese language of Diu*. Amsterdam: Landelijke Onderzoekschool Taalwetenschap.
- Clements, J. Clancy. 1996. *The genesis of a language: The formation and development of Korlai Portuguese*. Amsterdam: Benjamins.
- Clements, J. Clancy. 2009. *The legacy of Spanish and Portuguese: Colonial expansion and language change*. Cambridge: Cambridge University Press.
- Clements, J. Clancy. 2014. Form selection in contact languages: Evidence from some Portuguese- and Spanish-lexified contact varieties. In Patrícia Amaral & Ana Maria Carvalho (eds.), *Portuguese-Spanish interfaces: Diachrony, synchrony, and contact*, 377–401. Amsterdam & Philadelphia: John Benjamins.
- Clements, J. Clancy. 2018. The Iberian challenge: Creole languages beyond the plantation setting. *Journal of Pidgin and Creole Languages* 33(2). 447–452. DOI: 10.1075/jpcl.00025.cle.
- Clements, J. Clancy & Andrew Koontz-Garboden. 2002. Two Indo-Portuguese creoles in contrast. *Journal of Pidgin and Creole Languages* 17(2). 191–236.
- Clements, J. Clancy & Ana R. Luís. 2014. Contact intensity and the borrowing of bound morphology in Korlai Indo-Portuguese. In Francesco Gardani, Peter Arkadiev & Nino Amiridze (eds.), *Borrowed morphology*, 217–237. Boston & Berlin: De Gruyter Mouton.
- Davies, Mark. 2002. *Corpus del Español: 100 million words, 1200s–1900s*. <http://www.corpusdelespanol.org>.
- Davies, Mark & Michael Ferreira. 2006. *Corpus do Português: 45 million words, 1300s–1900s*. <http://www.corpusdoportugues.org>.
- Good, Jeff. 2012. Typologizing grammatical complexities, or why creoles may be paradigmatically simple but syntagmatically average. *Journal of Pidgin and Creole Languages* 27(1). 1–47.
- Klein, Wolfgang & Clive Perdue. 1992. *Utterance structure: Developing grammars again*. Amsterdam: Benjamins.
- Klein, Wolfgang & Clive Perdue. 1997. The basic variety (or: Couldn't natural languages be much simpler?) *Second Language Research* 13(4). 301–347.
- Kouwenberg, Silvia & Abigail Ramos-Michel. 2007. *Papiamentu (Creole Spanish/Portuguese)*. John Holm & Peter L. Patrick (eds.) (Westminster Creolistics Series 7). London: Battlebridge. 307–332.
- Lorenzino, Gerardo A. 2007. *Angolar (Creole Portuguese) or Lunga Ngola*. John Holm & Peter L. Patrick (eds.) (Westminster Creolistics Series 7). London: Battlebridge. 1–23.

- Maurer, Philippe. 1998. El papiamentu de Curazao. In Matthias Perl & Armin Schwegler (eds.), *América negra: Panorámica actual de los estudios lingüísticos sobre variedades hispanas, portuguesas y criollas*, 139–217. Frankfurt & Madrid: Vervuert.
- McColl Millar, Robert. 2007. *Trask's historical linguistics*. 2nd edn. London: Arnold.
- Mühlhäusler, Peter. 1997. *Pidgin & Creole linguistics*. 2nd edn. Colombo: Battlebridge Publications.
- Perdue, Clive. 1993a. *Adult language acquisition: Cross-linguistic perspectives. Field methods*, vol. 1. Cambridge: Cambridge University Press.
- Perdue, Clive. 1993b. *Adult language acquisition: Cross-linguistic perspectives. The results*, vol. 2. Cambridge: Cambridge University Press.
- Pinharanda, Mário. 2010. *Estudo da expressão morfo-sintática das categorias de tempo, modo e aspecto em Maquista*. University of Macau. (Doctoral dissertation).
- Schwegler, Armin. 1998. El Palenquero. In Matthias Perl & Armin Schwegler (eds.), *América negra: Panoramic actual de los estudios lingüísticos sobre variedades hispanas, portuguesas y criollas*, 218–291. Frankfurt: Vervuert.
- Schwegler, Armin & Kate Green. 2007. *Palenquero (Creole Spanish)* (Westminster Creolistics Series 7). London: Battlebridge. 273–306.
- Wohlgemuth, Jan. 2009. *A typology of verbal borrowings*, vol. 211. Berlin: Walter de Gruyter.

Appendix A Verbal paradigms in three Northern Indo-Portuguese Creoles (Clements & Koontz-Garboden 2002 and Cardoso 2009)

| | Korlai | Daman | Diu |
|------------------------|----------------------------------|--------------------------------|--------------------|
| Pres. -a | — | — | fal say |
| Pres./Inf. -a | halá say | fálá say | falá say |
| Pres. -e | — | — | beb drink |
| Pres./Inf. -e | bebé drink | bebé drink | bebé drink |
| Pres. -i | — | — | durm sleep |
| Pres./Inf. -i | drumí sleep | durmí sleep | durmí sleep |
| Pres./Inf. -u | tapú heat | babrú mutter | — |
| pres./past progressive | | | adverbial gerund |
| Pres. Ptcpl. -a | (te/ti) halán (AUX) saying | te/tiŋ fálán AUX saying | falán saying |
| -e | (te/ti) bebén (AUX) drinking | te/tiŋ bebén AUX drinking | bebén drinking |
| -i | (te/ti) drumín (AUX) sleeping | te/tiŋ durmín AUX sleeping | durmín sleeping |
| -u | (te/ti) tapún (AUX) heating | te/tiŋ babrún AUX grumbling | — |
| past | | | past |
| Past -a | haló said | fəló said | faló said |
| -e | bebéw drank | bebéw drank | bebéw drank |
| -i | drumíw slept | durmiw slept | durmíw slept |
| -u | tapú heated | babrú grumbled | — |
| pres./past pfct | | | participial adj. |
| Past Ptcpl. -a | (te/ti) halád (AUX) said | te/tiŋ fálád AUX said | falád said |
| -e | (te/ti) bebíd (AUX) drunk | te/tiŋ bebíd AUX drunk | bebíd drunk |
| -i | (te/ti) drumíd (AUX) slept | te/tiŋ durmíd AUX slept | durmíd slept |
| -u | (te/ti) tapúd (AUX) heated | te/tiŋ babrúd AUX grumble | — |

Appendix B The paradigms (partial) of the Portuguese verb system

Table 13: Portuguese verbal paradigms simple forms

| PRET | IMPFT | PRS | FUT | COND |
|------------|-------------|----------------|--------------|-------------|
| cantei | cantava | canto | cantarei | cantaria |
| cantaste | cantavas | cantas | cantarás | cantarias |
| cantou | cantava | canta | cantará | cantaria |
| cantámos | cantávamos | cantamos | cantaremos | cantaríamos |
| cantastes | cantáveis | cantais | cantareis | cantaríeis |
| cantaram | cantavam | cantam | cantarão | cantariam |
| PLUPERFECT | IMPFT SBJV | PRS SBJV/IMP | PERSONAL INF | |
| cantara | cantasse | cante | cantar | |
| cantaras | cantasses | cantes/canta | cantares | |
| cantara | cantasse | cante/cante | cantar | |
| cantáramos | cantássemos | cantemos | cantarmos | |
| cantáreis | cantásseis | canteis/cantai | cantardes | |
| cantaram | cantassem | cantem | cantarem | |
| PRET | IMPFT | PRS | FUT | COND |
| bebi | cantava | bebia | beberei | beberia |
| bebeste | cantavas | bebias | beberás | beberias |
| bebeu | cantava | bebia | beberá | beberia |
| bebemos | cantávamos | bebíamos | beberemos | beberíamos |
| bebesteis | cantáveis | bebíeis | bebereis | beberíeis |
| beberam | cantavam | bebiam | beberão | beberiam |
| PLUPERFECT | IMPFT SBJV | PRS SBJV/IMP | PERSONAL INF | |
| bebera | bebesse | beba | beber | |
| beberas | bebesse | bebas/bebe | beberes | |
| bebera | bebesse | beba/beba | beber | |
| bebêramos | bebêssemos | bebamos | bebermos | |
| bebêreis | bebêsseis | bebais/bebei | beberdes | |
| beberam | bebessem | bebam | beberem | |

Table 14: Portuguese verbal paradigms compound perfect forms

| IMPFT | PRS | FUT | COND |
|--------------------|------------------|-----------------|------------------|
| tinha cantado | tenho cantado | terei cantado | teria cantado |
| tinhas cantado | tens cantado | terás cantado | terias cantado |
| tinha cantado | tem cantado | terá cantado | teria cantado |
| tínhamos cantado | temos cantado | teremos cantado | teríamos cantado |
| tínheis cantado | tendes cantado | tereis cantado | teríeis cantado |
| tinham cantado | têm cantado | terão cantado | teriam cantado |
| IMPFT SBJV | PRS SBJV | PERSONAL INF | |
| tivesse cantado | tenha cantado | ter cantado | |
| tivesses cantado | tenhas cantado | teres cantado | |
| tivesse cantado | tenha cantado | ter cantado | |
| tivéssemos cantado | tenhamos cantado | termos cantado | |
| tivésseis cantado | tenhais cantado | terdes cantado | |
| tivessem cantado | tenham cantado | terem cantado | |
| IMPFT | PRS | FUT | COND |
| tinha bebido | tenho bebido | terei bebido | teria bebido |
| tinhas bebido | tens bebido | terás bebido | terias bebido |
| tinha bebido | tem bebido | terá bebido | teria bebido |
| tínhamos bebido | temos bebido | teremos bebido | teríamos bebido |
| tínheis bebido | tendes bebido | tereis bebido | teríeis bebido |
| tinham bebido | têm bebido | terão bebido | teriam bebido |
| IMPFT SBJV | PRS SBJV | PERSONAL INF | |
| tivesse bebido | tenha bebido | ter bebido | |
| tivesses bebido | tenhas bebido | teres bebido | |
| tivesse bebido | tenha bebido | ter bebido | |
| tivéssemos bebido | tenhamos bebido | termos bebido | |
| tivésseis bebido | tenhais bebido | terdes bebido | |
| tivessem bebido | tenham bebido | terem bebido | |